



**UNIVERSIDADE FEDERAL DA BAHIA  
INSTITUTO DE FÍSICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA**

**Rosana de Andrade Sousa**

# **Método de Sincronização de Redes Multicamadas Variantes no Tempo**

**Salvador-Bahia  
22 de abril de 2022**

Rosana de Andrade Sousa

## **Método de Sincronização de Redes Multicamadas Variantes no Tempo**

Proposta de tese apresentada ao Programa de Pós-graduação em Física da Universidade Federal da Bahia, como requisito parcial para obtenção do grau de doutora em Física.

Universidade Federal da Bahia – UFBA

Instituto de Física

Programa de Pós-Graduação em Física

Orientador: Dr. José Garcia Vivas Miranda

Coorientador: Dr. Raphael Silva do Rosário

Salvador-Bahia

22 de abril de 2022

Inevitavelmente, dedico esse trabalho a minha mãe, luz da minha vida. O seu amor incondicional, força, paz e sorriso foram as minhas fontes de energia para continuar após a sua partida. Obrigada por ter sido refúgio, calma, fortaleza, apoio, incentivo, esperança e muito amor. Guia-me de onde estiver.

# Agradecimentos

Agradeço imensamente a todos que me acompanharam durante o doutorado e, de alguma forma, contribuíram para o meu desenvolvimento pessoal ou acadêmico, tornando esse processo menos solitário.

Primeiramente, agradeço às pessoas mais importantes da minha vida, minha mãe Dolores Alves de Andrade e meu pai Manuel Francisco de Sousa. Agradeço por todos os sacrifícios que fizeram para que eu pudesse trilhar as minhas escolhas e, por terem sido fontes infindáveis de amor, incentivo e proteção. Infelizmente, vocês não puderam esperar este dia, mas de onde estiverem, recebam o meu amor e meu eterno agradecimento.

Agradeço ao meu orientador José Garcia Vivas Miranda, por sempre ter um olhar otimista e uma palavra de incentivo que, por diversas vezes, me convenceu a continuar. Do mesmo modo, agradeço ao meu coorientador Raphael Silva do Rosário por sempre encarar os desafios com entusiasmo e paciência. A vocês, agradeço a orientação, confiança, dedicação e leveza com que me conduziram nesse processo intrinsecamente pesado.

Agradeço o apoio e o amor dos meus irmãos Marlyps de Sousa, Robson Andrade e Rosângela Andrade e minha cunhada Geysa Mascarenhas. Agradeço, especialmente a Ró pelo companheirismo e dedicação. Por enfrentarmos juntas todos os desafios que a vida nos traz. Venceremos juntas! Agradeço às minhas sobrinhas Lara Beatriz, Ruth Maria e Manuella Andrade pelo amor e por entenderem a minha ausência.

Agradeço aos meus queridos amigos Sandra Oliveira e a Marcos Santos por sempre me acolherem com tanto amor e empatia. E, por estarem comigo em todos os momentos!

Agradeço ao meu amigo-irmão Rone Lemos por sempre ouvir as minhas inseguranças e anseios. E, por sempre tentar me fazer acreditar que as coisas dariam certo. Obrigada!

Agradeço aos meus amigos que estão fisicamente distantes. Talvez vocês não saibam o quanto pequenos gestos como uma mensagem aleatória ou uma ligação inesperada fizeram muita diferença nos meus dias. Agradeço principalmente a Chirlei Dias, Greicekelly Gusmão, Dante Ferraz, Mônia Macedo, Andrielle Prado, Wadna Nolasco, Samara Campos, Tam Alves, Isis Rolim, Sheila Freitas, Maurício Leite e Kel Novais.

Agradeço aos amigos do grupo de pesquisa, especialmente a Ana Paula Quixadá, Marina Souza, Tandara Benevides, Thaise Toutain, Caíque Moura, Vitor Sotero, Ângelo Frederico, Pietro Rebouças e Marcelo Cunha por todo conhecimento compartilhado. Sentirei falta das conversas aleatórias, das festinhas surpresas e dos cafezinhos da tarde no LaBios.

A Elaine Cambuí agradeço a amizade, carinho e apoio. Todas as suas contribuições nos meus trabalhos foram essenciais para o meu desenvolvimento como pesquisadora.

A Pedro Teodoro obrigada por sempre trazer sorrisos e leveza para a minha vida. Obrigada pela amizade sem medidas e, por estar sempre presente, mesmo morando distante.

A Raíssa Fernandes e Eslaine Santos sou igualmente grata pela ajuda, apoio, trocas, amizade e carinho construídos ao longo desses anos. Contem sempre comigo!

Aos amigos do Instituto de Física sou grata pelos momentos de partilha das alegrias, das angústias, das conquistas, anseios e resenhas. Agradeço especialmente a Manuela Arruda, Leonardo Fabio, Alfredo Blanco, Edwin Mozo, Caio Guimarães, Ademir Santos, Elenilson Nery, Eric Pinto, Patrícia Hepp, Felipe Ventura e Leonardo Cerqueira.

Agradeço a minha amiga-irmã Mariana Medeiros. Foram 7 anos compartilhando não só um lar, mas também muita amizade, carinho, respeito e cumplicidade. Poder dividir as minhas alegrias e frustrações com você e Andréia Simões foi essencial e tornou esse caminho mais leve. Obrigada, meninas! Quero levar a nossa amizade por toda a vida!

A Vinícius Nonato agradeço a amizade partilhada durante grande parte da minha vida acadêmica. Ela representou uma fonte enorme de cuidado, apoio, paciência e incentivo. Agradeço, sobretudo, pelas inúmeras tentativas de me fazer enxergar o melhor lado das coisas, das pessoas e de mim mesma. Obrigada!

A Vinícius Mendonça obrigada por estar sempre ao meu lado para me ouvir, apoiar e aconselhar tanto nos tempos tranquilos, como nos tempestivos. Obrigada por tudo!

A João Ricardo agradeço pela amizade. As nossas conversas que transitavam entre o doutorado, política, planos e resenhas, tornaram o isolamento menos pesado.

A Wallas Nascimento agradeço o carinho, a amizade e os cafés aos finais das tardes. Cafés estes, regados das nossas inseguranças acadêmicas e de muitas risadas.

A Marcos Paulo e Olívia Maria agradeço por me deixarem fazer parte da vida de vocês e, pela oportunidade de ter Elis Oliveira em minha vida. Obrigada por tudo!

A Romero Almeida e Emmanuel Cristiano agradeço por nossa amizade leve, sincera e acolhedora que, por muitas vezes, me ajudou a encarar as pressões do doutorado.

Agradeço aos funcionários do Instituto pelo acolhimento, orientações e disposição durante esses anos. Especialmente a Cristina Oliveira por tornar os almoços mais leves!

Agradeço aos professores que contribuíram para a minha formação, especialmente ao professor Frederico Vasconcelos, por ser um exemplo de humildade e humanidade.

A Paula Quintela obrigada por ter me acolhido com tanta dedicação, carinho e empatia. O seu apoio foi fundamental para mim. Serei sempre grata por tudo!

Por fim, agradeço a um Deus que eu não sei nomear ou definir, mas que sempre coloca pessoas maravilhosas em minha vida e que sempre me dá forças pra recomeçar.

A CAPES pelo apoio financeiro.

# Lista de abreviaturas e siglas

CHIRPS	Precipitação infravermelha do grupo de riscos climáticos com dados da estação
GEE	Google Earth Engine
GVT	Grafos Variantes no Tempo
IBGE	Instituto Brasileiro de Geografia e Estatística
MODIS	Espectrorradiômetro de Imagem de Resolução Moderada
NDVI	Índice de Vegetação por Diferença Normalizada
REA	Rede Estática Agregada

# Lista de tabelas

Tabela 1 – Matriz de adjacência do grafo da Figura 1 . . . . .	4
Tabela 3 – Operações realizadas nos dados para transformá-los em medidas quinzenais. . . . .	39
Tabela 4 – Índices das redes dos dados de evapotranspiração e de uma rede similar construída com conexões aleatórias. Os índices $\langle k \rangle$ , $\langle k^\omega \rangle$ e $\langle C \rangle$ se referem ao grau médio, grau ponderado médio e coeficiente de aglomeração médio, respectivamente. Os intervalos entre parênteses indicam os valores mínimos e máximos para os índices correspondentes. . . . .	45
Tabela 5 – Índices das redes dos dados de NDVI e de uma rede similar construída com conexões aleatórias. Os índices $\langle k \rangle$ , $\langle k^\omega \rangle$ e $\langle C \rangle$ se referem ao grau médio, grau ponderado médio e coeficiente de aglomeração médio, respectivamente. Os intervalos entre parênteses indicam os valores mínimos e máximos para os índices correspondentes. . . . .	48
Tabela 6 – Índices das redes dos dados de precipitação e de uma rede similar construída com conexões aleatórias. Os índices $\langle k \rangle$ , $\langle k^\omega \rangle$ e $\langle C \rangle$ se referem ao grau médio, grau ponderado médio e coeficiente de aglomeração médio, respectivamente. Os intervalos entre parênteses indicam os valores mínimos e máximos para os índices correspondentes. . . . .	50
Tabela 7 – Índices das redes dos dados de temperatura e de uma rede similar construída com conexões aleatórias. Os índices $\langle k \rangle$ , $\langle k^\omega \rangle$ e $\langle C \rangle$ se referem ao grau médio, grau ponderado médio e coeficiente de aglomeração médio, respectivamente. Os intervalos entre parênteses indicam os valores mínimos e máximos para os índices correspondentes. . . . .	52

# Lista de ilustrações

Figura 1 – Grafo: vértices e arestas . . . . .	3
Figura 2 – Grafo com hub . . . . .	5
Figura 3 – Grafo composto por 4 vértices e cálculo do coeficiente de aglomeração do vértice $v_1$ . . . . .	6
Figura 6 – (a) Ilustração de um grafo $\mathcal{G}$ composto por 4 grafos e 5 vértices. (b) Rede estática agregada associada a $\mathcal{G}$ . . . . .	11
Figura 7 – Motifs, lag e ordem . . . . .	13
Figura 8 – Método dos Motifs . . . . .	13
Figura 9 – Rede composta por 2 camadas com 5 vértices cada. . . . .	16
Figura 10 – Rede multicamadas variante no tempo com $P = 2$ e $\Gamma = 5$ . . . . .	28
Figura 11 – (a) Rede multicamadas com 5 vértices em cada camada e tempo de vida dos GVT igual a 4. (b) Rede de Incidência para os GVT considerados. . . . .	28
Figura 12 – (a) Rede multicamadas com 5 vértices em cada camada e tempo de vida dos GVT igual a 5. (b) Rede de Fidelidade para os GVT considerados. . . . .	29
Figura 13 – Rede de Incidência-Fidelidade para a rede multicamadas dos exemplos anteriores. . . . .	30
Figura 14 – Simulações para redes com duas camadas e diferentes tempos de vida dos GVT ao considerar um total de 100 amostras. Na construção dos GVT aleatórios a probabilidade da camada $G$ é igual a 0,4 e da camada $F$ é igual a 0,8. Valor médio dos índices (a) Incidência, (b) Fidelidade e (c) Incidência-Fidelidade. . . . .	32
Figura 15 – Índices de sincronização intercamadas para diferentes valores do acoplamento . . . . .	36
Figura 16 – Mapa da Bahia com seus respectivos biomas sendo representados por diferentes cores: em amarelo claro, o Cerrado; em laranja a Caatinga e em verde a Mata Atlântica . . . . .	37
Figura 17 – Fluxograma detalhando as etapas seguidas na aplicação do nosso método com os dados de variáveis climáticas e as ferramentas utilizadas. . . . .	38
Figura 18 – Séries temporais das variáveis climáticas do ano de 2017. Foram consideradas as seguintes cidades: Salvador, pertencente ao bioma Mata Atlântica (em verde); Barreiras, pertencente ao Cerrado (em preto) e Canudos, cujo bioma é a Caatinga (em vermelho). . . . .	40
Figura 19 – (a) Padrões de motifs adicionados ao método de sincronização por motifs.(b) Padrão de motif com 3 medidas seguidas iguais a zero. . . . .	41



Figura 20 – Vetores com 392 motifs de uma mesma variável para cidades diferentes e dividido em janelas ( $L_m$ ) de 12 motifs cada. Comparações realizadas para tempo de atraso $\tau = 0$ . . . . .	42
Figura 21 – Número de arestas dos GVT de acordo com o valor do $\tau$ para cada variável climática. . . . .	43
Figura 22 – REA das arestas hubs dos dados de evapotranspiração. Essa rede está sobreposta ao mapa do estado da Bahia, com as devidas especificações dos biomas presentes: Cerrado (lima), Caatinga (amarelo claro), Mata Atlântica (verde). . . . .	45
Figura 23 – REA das arestas hubs dos dados de evapotranspiração sobreposta ao mapa da cobertura do solo do estado da Bahia. . . . .	46
Figura 24 – REA das arestas hubs dos dados de NDVI. Esta rede está sobreposta ao mapa do estado da Bahia, com as devidas especificações dos biomas presentes: Cerrado (lima), Caatinga (amarelo claro), Mata Atlântica (verde). . . . .	47
Figura 25 – REA das arestas hubs dos dados de NDVI sobreposta ao mapa da cobertura do solo do estado da Bahia. . . . .	49
Figura 26 – REA das arestas hubs dos dados de precipitação. Esta rede está sobreposta ao mapa do estado da Bahia, com as devidas especificações dos biomas presentes: Cerrado (lima), Caatinga (amarelo claro), Mata Atlântica (verde). . . . .	50
Figura 27 – REA das arestas hubs dos dados de temperatura. Essa rede está sobreposta ao mapa do estado da Bahia, com as devidas especificações dos biomas presentes: Cerrado (lima), Caatinga (amarelo claro), Mata Atlântica (verde). . . . .	51
Figura 28 – Rede de Incidência-Fidelidade do par evapotranspiração - NDVI. . . . .	54
Figura 29 – Rede de Incidência-Fidelidade do par evapotranspiração - temperatura. . . . .	55
Figura 30 – Rede de Incidência-Fidelidade do par evapotranspiração - precipitação. . . . .	57
Figura 31 – Rede de Incidência-Fidelidade o par NDVI - precipitação. . . . .	57
Figura 32 – Exemplo de um motif construído com $\lambda = 1$ . . . . .	62
Figura 33 – Exemplo ilustrativo de uma série temporal de um período de 1 ano e 5 meses. . . . .	62
Figura 34 – (a) Série temporal reescrita como um vetor de motifs. (b) Exemplo do movimento das janelas para <i>Lag window</i> = 1. . . . .	63

# Sumário

<b>1</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>2</b>
1.1	Teoria dos Grafos	2
1.2	Redes Dinâmicas	7
1.3	Método de Sincronização por Motifs	12
1.4	Redes Multicamadas	15
1.5	Índices Incidência, Fidelidade, Incidência-Fidelidade para redes de palavras	19
1.5.1	Incidência	20
1.5.2	Fidelidade	20
1.5.3	Incidência-Fidelidade	21
<b>2</b>	<b>SINCRONIZAÇÃO EM REDES MULTICAMADAS VARIANTES NO TEMPO</b>	<b>22</b>
2.1	Introdução	24
2.2	O Método	27
2.3	Validações	30
2.3.1	Sincronização em Redes Multicamadas Aleatórias Variantes no Tempo	31
2.3.2	Oscilador de Rössler Acoplado	33
2.4	Aplicação	36
2.4.1	Materiais e Métodos	37
2.4.2	Resultados	42
2.4.2.1	Redes das Variáveis Individuais	42
2.4.2.2	Redes de Sincronização para Pares de Variáveis Climáticas	53
2.5	Conclusões	59
	<b>APÊNDICES</b>	<b>61</b>
	<b>APÊNDICE A – CONSTRUÇÃO DAS REDES UTILIZANDO O MÉTODO DE SINCRONIZAÇÃO POR MOTIFS</b>	<b>62</b>
	<b>APÊNDICE B – BIOMAS PRESENTES NO ESTADO DA BAHIA</b>	<b>64</b>
	<b>REFERÊNCIAS</b>	<b>67</b>

# Apresentação

Neste trabalho propomos um método capaz de quantificar a sincronização entre dois parâmetros distintos de um sistema que varia no tempo e, que pode ser descrito utilizando redes multicamadas. Nesse método, consideramos as conexões como um fator essencial na obtenção da medida de sincronização. Sendo essa avaliada a partir de duas perspectivas: a coocorrência das conexões de diferentes parâmetros em relação ao tempo e, em relação a assiduidade das conexões.

A construção deste método é embasada no conceito de redes, assim no primeiro capítulo desta tese abordamos uma breve introdução sobre a teoria dos grafos e seus índices básicos de análise de um sistema. Os sistemas considerados aqui exibem mudanças com o decorrer do tempo, e são avaliados a partir de diferentes parâmetros. Em vista disso, os conceitos de grafos variantes no tempo - GVT [1] e redes multicamadas [2–4] também são discutidos nesse capítulo.

Apresentamos ainda no Capítulo 1 uma descrição do método de sincronização por motifs [5] desenvolvido por Rosário e colaboradores, que transforma séries temporais em redes variantes no tempo, já que este método será utilizado no nosso trabalho. E, para finalizar o capítulo, apresentamos os índices de Incidência e Fidelidade [6] desenvolvidos no estudo de redes de palavras e que serviram como base para a concepção da nossa proposta.

O Capítulo 2 obedece o formato de artigo e, sendo assim, a primeira parte é composta por: resumo, palavras-chave e introdução. Nas seções 2.2 e 2.3, apresentamos a nossa metodologia e as validações desenvolvidas. A sua aplicação é discutida na seção seguinte, onde consideramos a sincronização intercamadas de diferentes combinações de 4 variáveis climáticas das 417 cidades do estado da Bahia em um período de 17 anos. Na última seção do capítulo, seção 2.5, apresentamos as conclusões finais sobre o método desenvolvido e sobre a sua aplicação.

Por fim, temos dois apêndices. No primeiro, justificamos a escolha dos valores dos parâmetros do método de sincronização por motifs [5]. E, no segundo, discorremos de forma simplificada sobre os biomas encontrados no estado da Bahia.

# 1 Fundamentação Teórica

Neste capítulo é apresentada a fundamentação teórica necessária para o trabalho desenvolvido. Essa fundamentação está distribuída ao longo de quatro seções: na primeira, apresentamos um breve compilado sobre a teoria dos grafos e alguns de seus índices de caracterização, juntamente com o conceito de grafos variantes no tempo e a sua importância no estudo de sistemas que apresentam uma dinâmica temporal. Na seção seguinte, mostramos o método de sincronização por motivos que visa construir redes a partir de séries temporais de um sistema qualquer. A terceira seção, traz uma discussão geral sobre um tipo de redes denominada de redes multicamadas, abordando brevemente sobre o seu formalismo, aplicações e sobre a sua importância na modelagem de sistemas. E na última seção, apresentamos os índices de caracterização das redes semânticas, denominados como: Incidência, Fidelidade e Incidência-Fidelidade, que são as bases para a construção dos índices de sincronização intercamadas propostos nesse trabalho.

## 1.1 Teoria dos Grafos

O termo *redes* já faz parte do cotidiano das pessoas. Desde a década de 90 com o surgimento das primeiras redes sociais, esse tipo de rede vem ganhando popularidade e conquistando novos usuários em todo mundo. Além das redes sociais, podemos citar outros tipos de redes que também nos cercam: redes de supermercado, redes de lojas de departamento, redes de computadores, redes de telefonia, redes de energia elétrica e redes de internet. Todos esses são exemplos de redes, mas, em linhas gerais, o que são as redes? Como surgiram? Como podemos caracterizá-las e utilizá-las para analisar sistemas das mais diversas áreas do conhecimento?

Uma rede pode ser definida como um conjunto de componentes que se conectam obedecendo a algum padrão de interação. As características destes componentes servem para classificar o tipo de rede que está sendo observada. Por exemplo, uma rede social é aquela composta por indivíduos que se conectam devido a algum critério de aproximação, como relações familiares, de trabalho, de amizade ou interesses em comuns. Já as redes biológicas descrevem os padrões de interação entre determinados elementos biológicos. Na Neurociência, por exemplo, utilizam-se as redes para representar padrões de conexões entre as células cerebrais. Já em Ecologia, as redes alimentares são utilizadas para quantificar a interação entre espécies diversas [7–9].

O estudo da redes teve início com o matemático suíço Leonhard Euler em 1736, ao desenvolver a *Teoria dos Grafos*. Entretanto, foi apenas no século 20 que o interesse da comunidade científica por essa área começou a se intensificar. Euler, a fim de solucionar

um problema famoso conhecido como o *problema das sete pontes de Königsberg*, abordou o problema descrevendo os seus constituintes utilizando uma linguagem matemática que recebeu o nome de *grafo*. Um grafo consiste numa composição de pontos, também chamados de vértices ou nós, e linhas, conhecidas como arestas, sendo uma forma de representar matematicamente uma rede.

Formalmente, um grafo  $G = (V, E)$  é composto por um conjunto finito de vértices  $V$ , e um conjunto de arestas  $E$ . Os elementos de  $V$  podem ser denotados como  $V = \{v_1, v_2, \dots, v_N\}$  onde  $N$  é o número total de vértices do grafo. Cada elemento do conjunto de arestas é definido por um par de vértices  $e_{i,j} = (v_i, v_j)$ , sendo  $v_i, v_j \in V$ .

A Figura 1 exibe um grafo composto por 5 vértices e 4 arestas. Nesse grafo a aresta  $e_{1,2}$ , por exemplo, é a que conecta o par de vértices  $(v_1, v_2)$ .

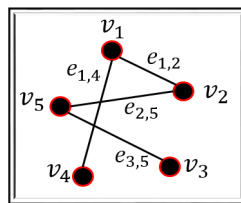


Figura 1 – Grafo: vértices e arestas

Os grafos são classificados de acordo com as suas características topológicas, sendo que no desenvolvimento desse trabalho os grafos utilizados foram os grafos simples e os grafos ponderados. Grafos ponderados são aqueles que possuem valores numéricos (pesos) associados às suas arestas. Tomemos como exemplo um grafo no qual os vértices correspondem as cidades e as arestas correspondem as rotas aéreas entre as cidades. Neste caso, o peso pode representar a quantidade de passageiros em cada viagem ou o número de vôos em um dia. Já os grafos simples são aqueles que: as arestas não possuem peso, não existe mais de uma aresta conectando os mesmos vértices e não há a presença de laços<sup>1</sup> [10, 11].

A teoria dos grafos dispõe de uma série de índices para caracterizar e analisar sistemas descritos como redes, tais como: grau dos vértices, grau ponderado, grau médio, coeficiente de aglomeração médio, modularidade. Estes índices, além de muitos outros, buscam uma melhor compreensão dos sistemas como um todo. As definições estabelecidas a seguir se encontram nas referências [7, 11–13].

Uma forma de quantificar quão conectado é um determinado vértice na rede é utilizando o índice *grau*. Para grafos simples, o grau de um vértice  $v_i$  é a soma do número de arestas conectadas a esse vértice. Matematicamente o grau  $k_i$  do vértice  $v_i$  é definido

<sup>1</sup> Uma aresta que conecta um vértice a ele mesmo.

	$v_1$	$v_2$	$v_3$	$v_4$	$v_5$
$v_1$	<b>0</b>	1	0	1	0
$v_2$	1	<b>0</b>	0	0	1
$v_3$	0	0	<b>0</b>	0	1
$v_4$	1	0	0	<b>0</b>	0
$v_5$	0	1	1	0	<b>0</b>

Tabela 1 – Matriz de adjacência do grafo da Figura 1

da seguinte forma:

$$k_i = \sum_{j=1}^N a_{i,j}, \quad (1.1)$$

onde  $N$  é o número total de vértices do grafo e  $a_{i,j}$  é o elemento de uma matriz denominada matriz de adjacência, a qual indica a presença das arestas no grafo. A matriz de adjacência é uma matriz quadrada  $A_{N \times N}$  de ordem  $N$  e o valor dos elementos  $a_{i,j}$  obedecem a seguinte condição:

$$a_{i,j} = \begin{cases} 1, & \text{se } v_i \text{ e } v_j \text{ estão conectados} \\ 0, & \text{caso contrário} \end{cases} \quad (1.2)$$

A matriz de adjacência para um grafo simples é simétrica e a sua diagonal principal é nula, uma vez que em grafos simples não existem laços. Para o grafo simples da Figura 1 podemos escrever a sua matriz de adjacência de acordo com a Tabela 1.

A matriz de adjacência para grafos ponderados é construída de forma similar. No entanto, o elemento  $a_{i,j}$  da matriz é preenchido com o peso  $w_{i,j}$  de cada aresta. Quando dois vértices não estão conectados, o elemento de matriz correspondente recebe o número 0. De acordo com a condição abaixo:

$$a_{i,j} = \begin{cases} w_{i,j}, & \text{onde } w_{i,j} \text{ é o peso da aresta entre o vértice } v_i \text{ e o vértice } v_j \\ 0, & \text{caso não haja aresta entre o vértice } v_i \text{ e o vértice } v_j \end{cases} \quad (1.3)$$

Ainda no contexto dos graus, a média aritmética dos graus de todos os vértices do grafo define o *grau médio*  $\langle k \rangle$  do grafo:

$$\langle k \rangle = \frac{\sum_{i=1}^N k_i}{N}, \quad (1.4)$$

onde  $N$  é o número total de vértices do grafo, como definido anteriormente.

Ao considerar grafos ponderados além do grau, pode-se calcular também o *grau ponderado* dos vértices. No cálculo desse índice, considera-se os pesos de todas as arestas conectadas a um determinado vértice  $v_i$ , conforme a equação abaixo:

$$k_i^w = \sum_{j=1}^N w_{i,j}, \quad (1.5)$$

onde  $w_{i,j}$  é o peso das arestas de acordo com a condição (1.3).

Além da informação sobre os graus individuais, a *distribuição dos graus* dos vértices do grafo também é de grande importância. O uso dessa distribuição auxilia no entendimento e na análise do comportamento da conectividade entre os elementos do sistema. A distribuição de graus é obtida por meio de um histograma de todos os possíveis graus  $k$  dos vértices e suas respectivas probabilidades  $p_k$ , sendo calculada da seguinte forma:

$$p_k = \frac{N_k}{N} \quad (1.6)$$

onde  $N_k$  é a quantidade de vértices com um dado grau  $k$ . A quantidade  $p_k$  expressa a probabilidade de um vértice  $v_i$ , escolhido aleatoriamente, ter grau igual a  $k$ .

Algumas distribuições, como as em forma de lei de potência, são melhores visualizadas quando se considera o gráfico da distribuição acumulada dos graus. Esse tipo de distribuição informa a probabilidade de encontrar vértices na rede com grau menor ou igual a um dado grau  $k$ .

O cálculo da distribuição acumulada considera o somatório de todas as probabilidades  $p_j$  dos vértices com grau menor ou igual a  $k$ , obtidas utilizando a Equação (1.6). Essa probabilidade é descrita como:

$$P_k = \sum_{j \leq k} p_j. \quad (1.7)$$

A distribuição de graus permite identificar a presença de vértices denominados *hubs*, esses vértices são definidos como vértices superconectados em relação aos demais vértices da rede [14]. São considerados hubs, os vértices cujo grau  $k_i$  obedece a condição  $k_i \gg \langle k \rangle + 2\sigma$ , onde  $\langle k \rangle$  é o grau médio da rede e  $\sigma$  o desvio padrão dos graus. O vértice  $v_1$  da Figura 2 é um exemplo de hub:

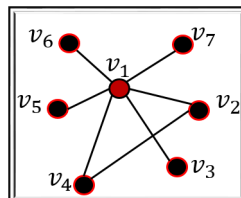


Figura 2 – Grafo com hub

Nos grafos cuja distribuição de graus segue uma distribuição gaussiana, pode-se inferir que os graus individuais se distanciam pouco do grau médio do grafo. Já os grafos cuja distribuição de graus seguem uma lei de potência, podemos assumir a existência de muitos vértices pouco conectados e poucos vértices muito conectados em relação aos demais vértices do grafo, os hubs [14].

Outro índice importante é o *coeficiente de aglomeração* de um vértice. Esse índice mede o nível de conexão entre os vértices do grafo. No seu cálculo considera-se a probabilidade de vértices adjacentes<sup>2</sup> a um dado vértice  $v_i$  sejam adjacentes entre si. Em seu cálculo considera-se a razão do número de arestas entre os vizinhos do vértice  $v_i$  dado por  $n_i$  e o maior número possível de arestas entre esses vizinhos. Assim, matematicamente, temos:

$$C_i = \frac{2n_i}{k_i(k_i - 1)}. \tag{1.8}$$

onde  $k_i$  é o grau do vértice  $v_i$ . O coeficiente de aglomeração do vértice  $v_1$  do grafo da Figura 3 é igual a  $\frac{1}{3}$ .

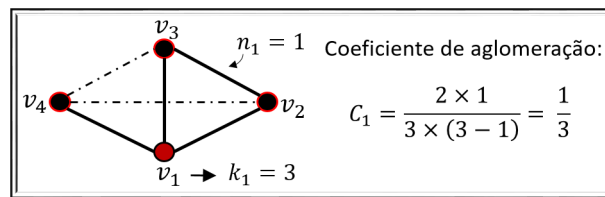


Figura 3 – Grafo composto por 4 vértices e cálculo do coeficiente de aglomeração do vértice  $v_1$ .

O coeficiente de aglomeração é um número no intervalo de 0 e 1 e é calculado somente em vértices com grau maior do que 1. Quando o coeficiente de aglomeração é nulo, significa que os vizinhos de um dado vértice  $v_i$  não estão conectados entre si. Por sua vez, quanto maior o número de conexões entre os vizinhos de um dado vértice  $v_i$ , maior será o coeficiente de aglomeração.

Para o grafo como um todo, define-se o *coeficiente de aglomeração médio* de um grafo,  $\langle C \rangle$ . Seu cálculo se dá pela média de todos coeficientes de aglomeração individuais, como definido na equação abaixo:

$$\langle C \rangle = \frac{\sum_{i=1}^N C_i}{N}. \tag{1.9}$$

É importante ressaltar que esta média considera apenas os vértices que possuem coeficiente de aglomeração definido.

Ao analisar uma rede é possível encontrar subgrupos de vértices, cuja densidade de arestas é maior dentro de cada subgrupo do que entre eles. Tais subgrupos são denominados *comunidades*. Por sua vez, uma *partição* é uma divisão específica de uma rede em comunidades, sendo a *modularidade* um conceito definido a partir de um modelo nulo e capaz de mensurar a qualidade de uma partição, isto é, determina se a partição é ou não significativa.

<sup>2</sup> Em um grafo simples  $G$ , dois vértices  $v_i$  e  $v_j$  são adjacentes caso exista uma aresta entre eles.



Considere um grafo  $G$  com  $m$  arestas e matriz de adjacência  $A$ , em que foi feita uma partição em comunidades  $G^*$ . A modularidade de  $G^*$  é dada por

$$Q(G^*) = \frac{1}{2m} \sum_i \sum_j \left( a_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j) \quad (1.10)$$

onde  $k_i$  é o grau do vértice  $i$  e  $\delta(c_i, c_j) = 1$  se  $v_i$  e  $v_j$  estiverem na mesma comunidade e 0, caso contrário. O termo  $\frac{k_i k_j}{2m}$  indica a probabilidade de existir uma aresta  $e_{ij}$  em um grafo aleatório, mantendo o grau dos vértices do grafo original.

Se o grafo for ponderado, a definição de modularidade de sua partição é dado por:

$$Q(G^*) = \frac{1}{2m'} \sum_i \sum_j \left( w_{ij} - \frac{k_i^w k_j^w}{2m'} \right) \delta(c_i, c_j) \quad (1.11)$$

onde  $m' = \sum_i k_i^w / 2$  e  $k_i^w$  é a definição para grau ponderado dada pela Equação 1.5.

Além desses, existem outros índices da teoria dos grafos que são utilizados para descrever e analisar sistemas que podem ser descritos como redes. No entanto, nos limitaremos apenas aos já apresentados, pois foram os utilizados no desenvolvimento desse trabalho.

## 1.2 Redes Dinâmicas

Como já sabemos, a teoria dos grafos é uma ferramenta importante para representar as relações entre os elementos dos sistemas estruturados como uma rede. Uma das formas de descrevê-los a partir da teoria dos grafos é utilizando o conceito de *rede estática*. Este tipo de rede considera todas as informações relevantes sobre o sistema colhidas dentro de um certo intervalo de tempo. Diversas redes reais, no entanto, são fortemente influenciadas pela dinâmica dos seus vértices e pelas mudanças das conexões entre eles.

O cérebro, por exemplo, é um sistema dinâmico onde novas sinapses são realizadas a todo momento; o índice pluviométrico das cidades pode ser avaliado para um dia apenas, mas também pode ser considerado um intervalo maior de tempo. As redes de relações sociais, por sua vez, podem mudar com a entrada ou saída de indivíduos. Assim, para estes sistemas é mais adequado descrevê-los utilizando as *redes dinâmicas*.

Diversos modelos surgiram com o objetivo de descrever as mudanças que ocorrem em um sistema. Albert e Barabasi [15], por exemplo, propuseram um modelo de construção de redes para descrever sistemas em que o número de vértices é fixo e as novas conexões acontecem preferencialmente entre os mesmos componentes. Esse modelo é clássico no estudo das redes e chama-se modelo de redes livres de escala. A seguir, veremos alguns modelos de construção dessas redes.

## Redes Temporais

De acordo com Holme e Saramäki [16], as redes que descrevem as modificações que ocorrem no sistema em um intervalo de tempo são chamadas de redes temporais. Essas redes podem ser representadas de duas formas, utilizando as sequências de contato ou os grafos de intervalo.

Nas sequências de contato, há um conjunto de  $N$  vértices  $V$  interagindo entre si em determinados momentos, e as durações das interações são desprezíveis. Um conjunto de contatos  $C$ , é dado por  $(i, j, t)$  onde  $i, j \in V$  e  $t$  denota o tempo. Temos ainda um conjunto de  $M$  arestas  $E$ , e, para  $e \in E$ , um conjunto não vazio de tempos de contatos  $T_e = \{t_1, t_2, \dots, t_n\}$ .

Alguns sistemas típicos que podem ser representados como uma sequência de contato incluem dados de comunicação, como: conjuntos de e-mails, telefonemas, mensagens de texto, etc [16]. A Figura 4(a) exemplifica uma sequência de contato composta por um conjunto de 4 vértices. As arestas dessa sequência expressam os pontos no tempo em que elas ocorrem, tanto numericamente, quanto graficamente, por meio de uma linha do tempo.

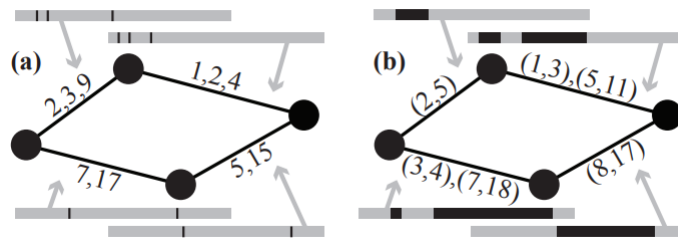


Figura 4 – (a) Sequência de contato e (b) Grafo de intervalo. As barras em cinza indicam as linhas do tempo e os momentos das conexões são marcados pelos campos em preto. Fonte: [16].

A segunda forma de representar as redes temporais é o grafo de intervalos. Neste grafo as arestas são ativas durante um intervalo de tempo, dessa forma para cada aresta  $e$  do grafo, temos o conjunto  $T_e = \{(t_1, t'_1), \dots, (t_n, t'_n)\}$ , onde os parênteses indicam os períodos de atividade da aresta.

Alguns sistemas que são naturais para modelar como grafos de intervalo, incluem: redes de proximidade; teias alimentares sazonais, onde um intervalo de tempo representa que uma espécie é a principal fonte de alimento de outra em alguma época do ano e sistemas de infraestrutura como a Internet [16].

A Figura 4(b) exemplifica um grafo de intervalo composto por um conjunto de 4 vértices. As arestas desse grafo expressam os intervalos de tempo em que essas arestas ocorrem. Temos ainda, as linhas do tempo dessas arestas, representando a sua presença por meio de pontos estendidos.

## Grafos Temporais

Para descrever as conexões entre os componentes de um sistema no decorrer do tempo, Kostakos [17] construiu uma ferramenta denominada *grafo temporal*.

Em um grafo temporal o sistema é caracterizado como uma rede, onde os vértices são distribuídos em uma "linha do tempo" conforme o momento em que realizam conexões. As suas arestas apontam a orientação das conexões, podem ser ponderadas a depender do tipo de conexão cometida e, podem ser pontilhadas ou contínuas, a depender do momento em que as conexões são realizadas.

Para ilustrar a construção de um grafo temporal consideramos uma lista de e-mails trocados entre 5 pessoas em um intervalo de 20 dias como mostra a Tabela 2:

Emissor	Receptor	Tempo
A	B	$t_1 = 0$
A	C,E	$t_2 = 1$
E	D	$t_3 = 3$
B	C	$t_4 = 5$
B	D	$t_5 = 9$
D	B	$t_6 = 14$
A	D	$t_7 = 20$

Tabela 2 – Conjunto de dados de e-mail: lista de e-mails trocados entre pessoas em momentos distintos (dias). Fonte: Figura adaptada de [17].

Para construir o grafo temporal admitimos os seguintes passos:

1. Para cada instante de tempo  $t_x$  criamos os vértices que realizam alguma conexão naquele instante de tempo. Portanto, para o tempo  $t_1$  temos o conjunto  $\{At_1, Bt_1\}$ . Temos também que a pessoa  $A$  pode ser representada pelo conjunto de instâncias  $\{At_1, At_2, At_7\}$  que indica todos os tempos em que ela participa de alguma conexão.
2. Para cada conjunto de instâncias, ligamos pares consecutivos  $\{At_x, At_{x+1}\}$  com arestas direcionadas de peso  $t_{x+1} - t_x$ , representando a distância temporal entre o par. Por exemplo, o peso entre os vértices  $At_2$  e  $At_7$  é de 19 dias ( $20 - 1$ ).
3. Por fim, utilizamos arestas direcionadas não ponderadas para vincular instâncias de vértices que participam de uma transação de e-mail. Assim, um e-mail de  $A$  para  $B$  no momento  $t_x$  é instanciado como uma aresta direcionada e não ponderada entre  $At_x$  e  $Bt_x$ .

A Figura 5 apresenta o grafo temporal resultante da lista de trocas de e-mail da Tabela 2.

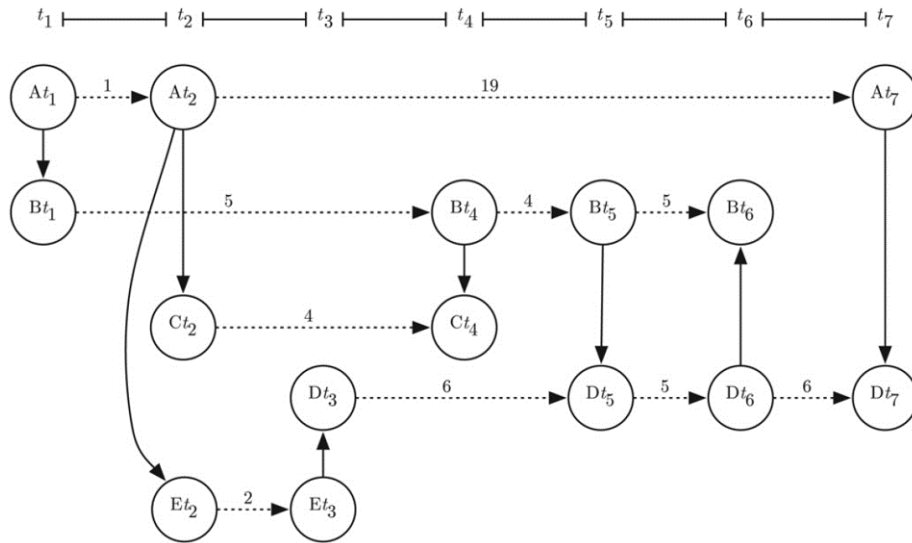


Figura 5 – Representação do grafo temporal da Tabela 2. Fonte: [17]

Para analisar esse grafo o autor definiu 5 métricas, sendo elas: proximidade temporal, proximidade temporal média, proximidade geodésica, proximidade geodésica média e disponibilidade temporal. Em síntese, essas métricas têm como objetivo quantificar a relação entre os vértices ao longo do tempo e o papel de cada vértice no contexto temporal de toda a rede.

### Grafos Variantes no Tempo - GVT

Outra forma de representar e estudar os efeitos do tempo em um sistema é utilizando os *Grafos Variantes no Tempo-GVT*, que são grafos em que as conexões aparecem e desaparecem ao longo do tempo entre um conjunto fixo de vértices. Nicosia e Tang [1] definiram um grafo variante no tempo como uma sequência ordenada  $\mathcal{G} = \{G_1, G_2, \dots, G_M\}$  composta por  $M$  grafos com  $N$  vértices. Nesta sequência, cada grafo  $G_s$  representa o estado da rede, no tempo  $t_s$ , onde  $s = 1, 2, 3, \dots, M$ .

Um grafo  $\mathcal{G}$  pode ser descrito como uma matriz de adjacência dependente do tempo  $A(t_s)$  onde  $a_{i,j}(t_s)$  são os elementos da matriz de adjacência do grafo  $G_s$  no instante  $t_s$  para  $s = 1, 2, 3, \dots, M$ .

$$a_{i,j}(t_s) = \begin{cases} 1, & \text{se } v_i \text{ e } v_j \text{ estão conectados} \\ 0, & \text{caso contrário} \end{cases} \quad (1.12)$$

A Figura 6(a) exemplifica um grafo variante no tempo composto por 4 grafos com 5 vértices, onde as diferentes configurações para cada instante de tempo são consequência

das representam das mudanças reais ocorridas no sistema. Para cada instante de tempo de  $\mathcal{G}$  é possível analisar o sistema utilizando os índices apresentados na seção 1.1.

Ao descartar a ordem temporal das arestas de cada grafo que compõe  $\mathcal{G}$  e, assim, considerar todas as conexões acontecendo ao mesmo tempo, obtemos a *Rede Estática Agregada-REA* [1]. A REA é um grafo ponderado que contém todas as conexões existentes em  $\mathcal{G}$  durante o intervalo de tempo  $[t_1, t_M]$ .

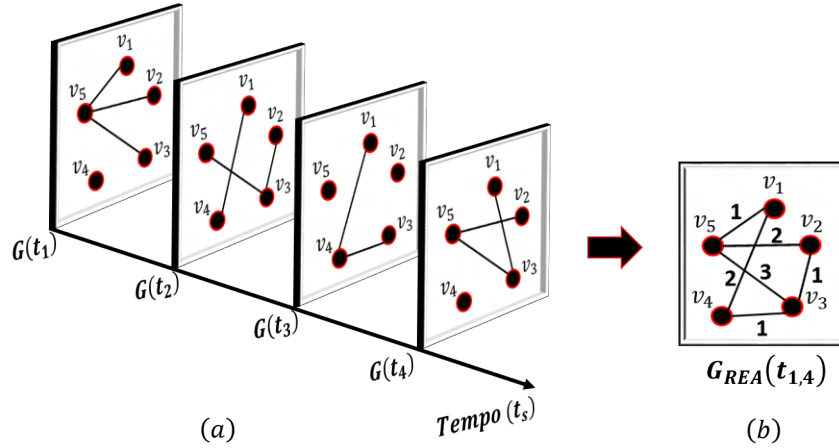


Figura 6 – (a) Ilustração de um grafo  $\mathcal{G}$  composto por 4 grafos e 5 vértices. (b) Rede estática agregada associada a  $\mathcal{G}$ .

A Figura 6(b) exibe a REA ao considerar  $\mathcal{G}$  no intervalo de  $t_1$  a  $t_4$ . Observe que a REA,  $G_{REA}(t_{1,4})$ , é um grafo ponderado em que os pesos das arestas indicam a frequência que determinada aresta aconteceu.

Casteigts e colaboradores [18] formalizaram matematicamente diversos conceitos relativos aos estudos das redes dinâmicas e definiram os GVT como uma quintupla  $\mathcal{G} = (V, E, \Gamma, \rho, \varsigma)$ , onde:

- $V$  representa o conjunto fixo de vértices de  $\mathcal{G}$ ;
- $E \subseteq V \times V$  representa o conjunto de arestas de  $\mathcal{G}$ ;
- $\Gamma$  é o tempo de vida do sistema;
- $\rho : E \times \varsigma \rightarrow \{0, 1\}$  é a função de presença que garante a existência de uma dada aresta em um dado instante de tempo;
- $\varsigma : E \times \Gamma \rightarrow \mathbb{N}$  é a função latência que diz respeito ao tempo necessário para estabelecer o relacionamento entre dois vértices em um dado instante.

É possível ainda definir uma função presença de vértice, indicando que o número de vértices conectados muda com o decorrer do tempo. No exemplo da Figura 6, o tempo

de vida de  $\mathcal{G}$  é igual a 4. Utilizaremos esta definição para nos referir ao tempo  $t_M$  definido por Nicosia e Tang [1].

Os grafos dinâmicos têm os benefícios analíticos da análise dos grafos estáticos e, ao mesmo tempo, retêm todas as informações temporais que podem estar disponíveis para os vértices [17]. Embora bem estabelecido, existem trabalhos que trazem variações no formalismo dos GVT proposto por Casteigts e colaboradores. Por exemplo, nos GVT não existem auto-conexões de um vértice  $v_j$  pertencente ao tempo  $t_{s-1}$  para  $t_s$ , dessa forma assume-se que os vértices "guardam as informações" para uma possível difusão posterior. Baseados nisso, Wehmuth e colaboradores [19] construíram um modelo unificado para os GVT considerando explicitamente estas auto-conexões que seriam responsáveis para o armazenamento e difusão das informações. A partir desse modelo geral, seria possível retomar o modelo dos GVT formalizado por Casteigts e colaboradores [18].

Os GVT têm sido utilizados nas análises de redes e construções de modelos das mais diversas áreas da pesquisa científica. Sousa e colaboradores [20] desenvolveram um modelo de interação preferencial baseado nos pesos das arestas e utilizaram os GVT para representar a evolução dos pesos na construção da rede. Rosário e colaboradores [5] utilizaram os GVT para desenvolver um modelo que permite construir redes a partir da sincronização por motifs entre séries temporais e aplicaram esse modelo para analisar a conectividade funcional do cérebro humano. Cunha e colaboradores [21] usaram os GVT para construir redes de cliques de títulos de publicações científicas e calcularam a entropia dessas redes, avaliando os seus máximos e mínimos. Santoro e Quattrociocchi [22] utilizaram as redes dinâmicas na análise de redes sociais. Hugo Saba e colaboradores [23] construíram redes de correlação entre as ocorrências de casos de dengue entre cidades do estado da Bahia. Por sua vez, os GVT foram utilizados também para a elaboração do trabalho aqui apresentado.

### 1.3 Método de Sincronização por Motifs

O método Sincronização por Motifs [5, 24] permite construir redes a partir de séries temporais. Para isso o método decompõe as séries temporais como uma sucessão de padrões pré-definidos, denominados *motifs*, que, geralmente, são classificados como aclives, declives, picos e vales.

Os motifs são caracterizados de acordo com os índices definidos como ordem ( $n$ ) e lag ( $\lambda$ ). A ordem é o número de pontos usados para a construção dos motifs, com  $n!$  sendo o número dos motifs possíveis. Já o lag é o intervalo entre os pontos dos motifs.

A Figura 7(a) exibe uma série temporal considerando os casos em que o lag é igual a 1 ou 2. Para  $\lambda = 1$  os pontos da construção dos motifs são consecutivos, no entanto, se  $\lambda > 1$  é necessário considerar um intervalo de tamanho  $\lambda$  entre os pontos. A Figura

7(b), exemplifica um motif de ordem 3 e as 6 configurações diferentes para esse valor de  $n$ .

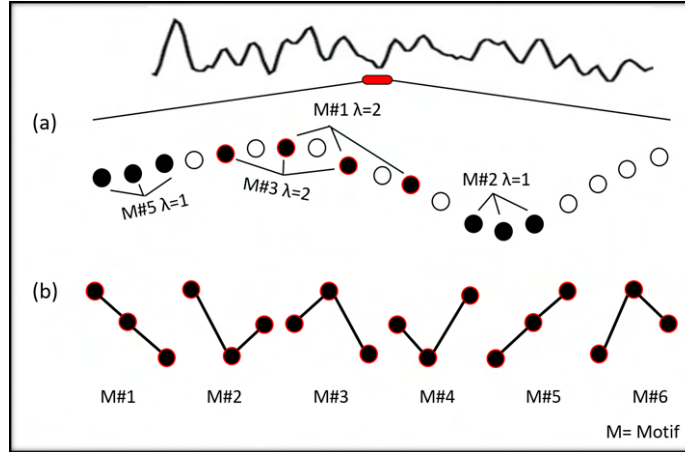


Figura 7 – (a) Série temporal transcrita como motivos com o lag igual a 1 ou 2. (b) Possíveis configurações para um motif de ordem 3. Fonte: Figura adaptada de [5, 24].

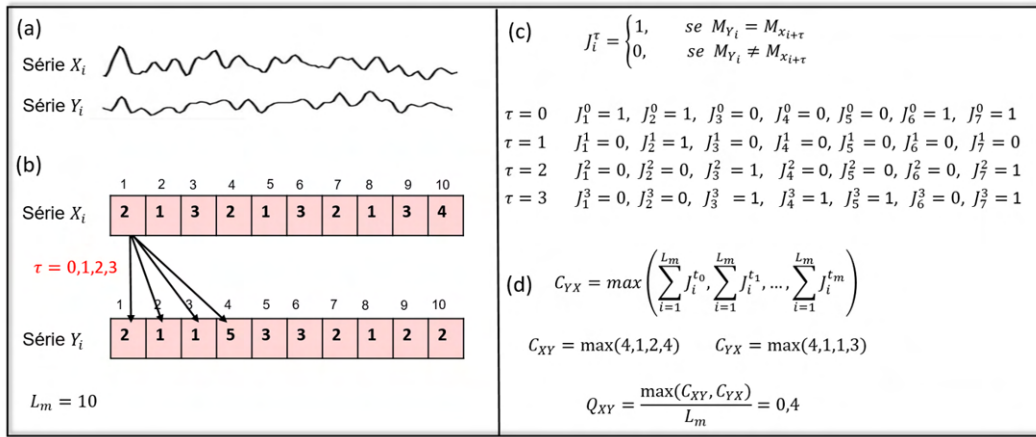


Figura 8 – Etapas da implementação do método de sincronização por motivos [5]. (a) Séries temporais  $X_i$  e  $Y_i$ . (b) Séries reescritas como uma sequência de motivos com tamanho  $L_m = 10$ , setas indicam o processo de cálculo da sincronização entre o primeiro motif da série  $X_i$  e motivos da série  $Y_i$  com tempo de atraso  $\tau = 0, 1, 2, 3$ . (c) Cálculo do índice de contagem  $J_i^\tau$ . (d) Cálculo do grau de correlação  $Q_{xy}$ . Fonte: Figura adaptada de [25].

O esquema da implementação do método de sincronização por motivos é detalhado a seguir e pode ser acompanhado na Figura 8.

1. Considere duas séries temporais  $X_i$  e  $Y_i$  e as reescreva como uma sequência de motivos. Por exemplo, as séries  $X_i$  e  $Y_i$  da Figura 8(a) foram reescritas como uma sequência de motivos  $M_{x_i}$  e  $M_{y_i}$ , da Figura 8(b). O parâmetro  $L_m$  é a janela do sinal e indica o tamanho da sequência (número de motivos). O parâmetro  $\tau$  é o tempo de atraso, e será utilizado na comparação entre as séries.
2. Compare estas séries para diferentes tempos de atraso  $\tau$ . A comparação entre duas séries temporais considera um valor fixo da janela e é realizada a partir da comparação

entre os elementos da sequência de motifs de cada série. Dessa forma é comparado o motif  $M_{x_i}$  que ocupa a posição  $i$  da série  $X_i$  com o motif  $M_{y_i}$  que ocupa a posição  $i$  da série  $Y_i$ , este processo é realizado para cada posição  $i$ . O tempo de atraso informa como deve ser feita a escolha dos elementos que serão comparados, como mostra a Figura 8(b).

Para medir o quanto as séries são semelhantes é definido um índice de contagem  $J_i^\tau$ , dado pela condição abaixo:

$$J_i^\tau = \begin{cases} 1, & \text{se } M_{x_i} = M_{y_{i+\tau}} \\ 0, & \text{caso contrário} \end{cases} \quad (1.13)$$

onde os valores de  $M_{x_i}$  e  $M_{y_i}$  correspondem aos motifs que ocupam a posição  $i$  nas sequências  $X_i$  e  $Y_i$ , para o tempo de atraso  $\tau$ . Na Figura 8(c) é calculado o valor de  $J_i^\tau$  para as sequências da Figura 8(b), para  $\tau$  de 0 à 3.

3. Calcula-se o número de vezes que o mesmo motif é encontrado em  $Y_i$  imediatamente após ter sido encontrado em  $X_i$  e obtém-se o valor de  $C_{xy}$ .

$$C_{xy} = \max \left( \sum_{i=1}^{L_m} J_i^{\tau_0}, \sum_{i=1}^{L_m} J_i^{\tau_1}, \dots, \sum_{i=1}^{L_m} J_i^{\tau_n} \right). \quad (1.14)$$

No exemplo, o valor máximo de  $C_{xy}$  e  $C_{yx}$  é igual a 4 (ver Figura 8(d)).

4. E por fim, é definido o grau de correlação  $Q_{xy}$  dado por

$$Q_{xy} = \frac{\max(C_{xy}, C_{yx})}{L_m}. \quad (1.15)$$

Dessa forma, de acordo com a Figura 8(d), têm-se que  $Q_{xy} = \frac{4}{10} = 0,4$ .

Após o grau de correlação  $Q_{xy}$  ser encontrado é necessário avaliar se o seu valor representa, de fato, uma medida de correlação ou se representa apenas uma conexão aleatória. Para isso, adota-se o seguinte critério:

- Para cada uma das séries temporais embaralha-se a sequência de motifs com o objetivo de trocar as posições dos motifs.
- Utiliza-se o método de sincronização por motifs na sequência embaralhada. Desse modo, os coeficientes  $Q_{xy}$  são definidos para todos os sinais aleatorizados. O próximo passo consiste em construir as redes sem considerar nenhum teste de significância.
- Escreve-se um vetor de arestas contendo os valores de  $Q_{xy}$  para todas as redes geradas. Esses valores são distribuídos em ordem crescente e, então obtém-se o valor de  $Q_{xy}$  correspondente a 95% do tamanho total deste vetor, que é definido como o *limiar*.



Após este processo as redes são construídas considerando a seguinte condição: se o grau de correlação  $Q_{xy}$  de uma aresta  $a_{xy}$  for maior ou igual ao valor do *limiar*, o elemento  $a_{xy}$  da matriz de adjacência dessa rede recebe o valor 1, caso contrário,  $a_{xy} \equiv 0$ . O *limiar* garante que as arestas da rede tenham 95% de chances de não serem frutos do acaso.

Por último, é possível construir a rede que representa o padrão de conectividade do sistema para uma dada janela. Ao deslocar a janela, todo o processo é repetido. O conjunto de redes que são geradas ao considerar todas as janelas formam um grafo variante no tempo  $\mathcal{G}$ . A soma de todas as conexões que foram realizadas durante em  $\mathcal{G}$  define a *REA*. A partir dela, calcula-se os graus dos vértices, o coeficiente de aglomeração, a distribuição de pesos das arestas, dentre outros.

## 1.4 Redes Multicamadas

Um conceito importante no estudo das redes é o de *redes multicamadas*. Embora seja uma abordagem relativamente nova, ela vem sendo bastante utilizada. Tratar os sistemas a partir de redes multicamadas permite novas compreensões sobre os seus mecanismos e os seus padrões de conexões. Mas, afinal, o que difere uma rede tradicional, como a que vimos até aqui, das redes multicamadas?

A maioria dos fenômenos que ocorrem ao nosso redor é o resultado da organização dinâmica emergente de sistemas compostos por uma diversidade de entidades que interagem entre si através de padrões não tão simples [2]. Neste contexto, as redes multicamadas buscam descrever explicitamente esses sistemas interligados através de diferentes categorias de conexões: cada canal (relacionamento, atividade, categoria, etc) é representado por uma camada e o mesmo vértice pode ter diferentes tipos de interações em cada camada [2].

Apesar de ser estudada a partir de uma abordagem tradicional, a rede social possui características que a tornam uma boa candidata ao estudo com multicamadas. Na abordagem tradicional, os diversos tipos de relacionamentos entre os indivíduos de uma rede social (amizade, proximidade, parentesco, etc) são tratados como conexões equivalentes, ou seja, não há distinção entre elas. Já na abordagem multicamadas, a ampla variedade de interações humanas pode ser categorizada em camadas diferentes e as análises são realizadas a partir da caracterização de cada camada. Essa forma de lidar com o sistema pode, possivelmente, alterar as conclusões sobre o mesmo [2, 3].

De acordo com Boccaletti e colaboradores [2], uma rede multicamadas  $\mathcal{M}$  pode ser definida como um par  $(g, c)$  onde  $g$  se refere ao conjunto de grafos da rede multicamadas e  $c$  ao conjunto de arestas existentes entre essas camadas e são chamadas de *arestas intercamadas* (do inglês *interlayer*).

Especificamente o conjunto  $g$  é dado por  $g = \{G_\alpha, \alpha \in \{1, 2, \dots, P\}\}$ , onde  $G_\alpha =$

$(V_\alpha, E_\alpha)$  corresponde ao grafo de uma camada qualquer  $\alpha$  e  $P$  é o número máximo de camadas. De forma análoga a definida na seção 1.1 o conjunto de vértices do grafo  $G_\alpha$  é dado por  $V_\alpha = \{v_1^\alpha, v_2^\alpha, \dots, v_{N_\alpha}^\alpha\}$  onde  $N_\alpha$  é o número total de vértices do grafo  $G_\alpha$  e  $E_\alpha$  é o seu conjunto de arestas. As arestas do conjunto  $E_\alpha$  são chamadas de *arestas intracamadas* (do inglês *intralayer*), que consistem nas arestas que ocorrem entre os vértices de um mesmo grafo  $G_\alpha$ . O conjunto de arestas intercamadas, citados anteriormente, é dado por  $c = \{E_{\alpha,\beta} \subseteq V_\alpha \times V_\beta; \alpha \text{ e } \beta \in \{1, 2, \dots, P\}, \alpha \neq \beta\}$ .

A Figura 9 ilustra uma rede com 2 camadas  $g = \{\alpha, \beta\}$ , em que  $\alpha = G_1$  e  $\beta = G_2$ , onde as arestas dentro de cada camada representam as conexões intracamadas e as arestas em vermelho as conexões intercamadas. Nessa rede, todas as camadas possuem um total de 5 vértices e as conexões intracamadas seguem diferentes padrões de conexão.

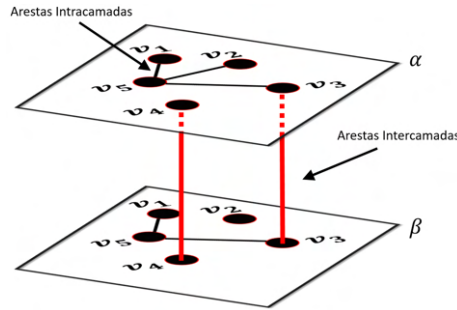


Figura 9 – Rede composta por 2 camadas com 5 vértices cada.

A matriz de adjacência para cada camada  $G_\alpha$ , pode ser denotada por  $A^{[\alpha]} = (a_{ij}^\alpha) \in \mathbb{R}^{N_\alpha \times N_\alpha}$  onde

$$a_{i,j}^\alpha = \begin{cases} 1, & \text{se } v_i^\alpha \text{ e } v_j^\alpha \in E_\alpha \\ 0, & \text{caso contrário} \end{cases} \quad (1.16)$$

para  $1 \leq i, j \leq N_\alpha$  e  $1 \leq \alpha \leq P$ .

No estudo das redes multicamadas é definida ainda a *matriz de adjacência intercamadas*, cujos elementos indicam a presença de arestas entre duas camadas  $\alpha$  e  $\beta$  quaisquer. Ela é dada por  $A^{[\alpha,\beta]} = (a_{i,j}^{\alpha,\beta}) \in \mathbb{R}^{N_\alpha \times N_\beta}$ , cujos valores dos elementos  $a_{i,j}^{\alpha,\beta}$  obedecem a condição:

$$a_{i,j}^{\alpha,\beta} = \begin{cases} 1, & \text{se } (v_i^\alpha, v_j^\beta) \in E_{\alpha,\beta} \\ 0, & \text{caso contrário} \end{cases} \quad (1.17)$$

A *matriz de adjacência intercamadas*  $A^{[\alpha,\beta]}$  da rede multicamadas da Figura 9 é dada por:

onde as linhas representam o conjunto de vértices da camada  $\alpha$  e as colunas o conjunto de vértices da camada  $\beta$ . Os critérios das conexões intercamadas dependem do

	$v_1^\beta$	$v_2^\beta$	$v_3^\beta$	$v_4^\beta$	$v_5^\beta$
$v_1^\alpha$	0	0	0	0	0
$v_2^\alpha$	0	0	0	0	0
$v_3^\alpha$	0	0	1	0	0
$v_4^\alpha$	0	0	0	1	0
$v_5^\alpha$	0	0	0	0	0

problema de pesquisa abordado. Alguns exemplos de trabalhos com redes multicamadas serão vistos ainda nessa seção.

Existe um tipo especial de redes multicamadas, chamadas de *redes multiplex*. Para este tipo de rede o conjunto de vértices é o mesmo em todas as camadas e as conexões intercamadas são permitidas apenas entre vértices equivalentes. Uma rede multiplex pode ser reescrita como uma supra matriz de adjacência. Essa matriz é composta de blocos que consideram todas as conexões da rede, tanto as intraconexões como as interconexões. Uma supra matriz de adjacência  $\tilde{A}$  possui dimensão  $\mathbb{R}^{N\alpha \times N\alpha}$  e é dada por:

$$\tilde{A} = \begin{pmatrix} A_1 & I_N & \cdots & I_N \\ I_N & A_2 & \cdots & I_N \\ \vdots & \vdots & \ddots & \vdots \\ I_N & I_N & \cdots & A_\alpha \end{pmatrix}$$

sendo  $A_\alpha$  a matriz de adjacência de cada camada  $\alpha$ ,  $I_N$  uma matriz identidade  $N$ -dimensional, onde  $N$  o número de vértices das camadas.

As redes variantes no tempo discutidas na seção 1.2, podem ser vistas como redes multicamadas ao considerar que cada camada se refere a um instante de tempo  $t_s$  da rede variante no tempo. Nessa representação o conjunto de arestas intercamadas é vazio. Contudo, podemos considerar ainda, sistemas descritos como redes multicamadas em que as conexões intracamadas mudam com o decorrer do tempo.

As redes adotadas para o desenvolvimento deste trabalho possuem o mesmo conjunto de vértices em todas as camadas, mas não há a presença de conexões intercamadas. Ressaltamos, ainda, que existe uma série de métricas na literatura das redes multicamadas, mas que elas não serão tratadas nessa discussão, por não serem utilizadas nesse trabalho. Maiores informações sobre estes tópicos podem ser encontradas nas referências [2–4].

De acordo com Kivelä e colaboradores [3], existem variadas formas de representar as redes multicamadas, como por exemplo: as camadas podem se conectar por meio de conexões intercamadas; todas as camadas podem conter os mesmos vértices ou número diferentes de vértices; as conexões intercamadas podem indicar a dependência entre as camadas; os vértices podem ser rotulados e conectados devido a alguma propriedade específica, entre outros. Ainda de acordo com o autor, são estas particularidades na estrutura de cada rede que vão classificar o seu tipo, como: redes de redes, redes heterogêneas, redes

multiplex, redes interdependentes, redes multiníveis, redes de vértices coloridos, dentre outros.

As redes multicamadas têm sido aplicadas para estudar desde a sociedade [26, 27] até a psicologia [28]. Sistemas que usualmente são tratados como uma rede tradicional, podem agora ser analisados sob outra perspectiva, um exemplo são as redes de citação. Geralmente nos trabalhos com esse tipo de rede são consideradas apenas uma camada [29] cuja construção desconsidera particularidades, tais como: a revista que o artigo foi publicado, se os artigos foram publicados na mesma conferência, etc. O uso das redes multicamadas abre novas possibilidades.

Zingg e colaboradores [30] analisaram a influência das relações sociais dos autores de um artigo científico na taxa de citação de novas publicações. Para isso consideraram dados de nove periódicos de física e utilizaram o conceito de redes multicamadas na análise desses dados. Foram consideradas redes de duas camadas: social e de publicações. Na camada social os autores são os vértices e as arestas são as relações de coautoria. A camada de publicação é composta pelos artigos que são os vértices e as citações entre artigos são as arestas. As relações intercamadas representam a relação entre autores e artigos. Eles encontraram uma tendência geral de artigos escritos por autores com mais coautores anteriores atraírem a atenção mais rapidamente, mas também serem esquecidos mais cedo. Esse efeito também é encontrado quando consideram o número de artigos anteriores dos autores ou quando controlam o tempo em que um artigo foi publicado.

Horvát e Zweig [31, 32] utilizaram uma estrutura multiplex para construir uma rede de indicação de filmes. Os dados usados foram disponibilizados pela Netflix e continham as classificações de filmes atribuídas por seus usuários. Na construção dessa rede, os vértices são os filmes e uma aresta é estabelecida quando um usuário avalia um par de filmes. Com relação as camadas, elas indicam a avaliação dos usuários para um par de filmes, que podem ser: positiva (quando ambos são apreciados), negativa (quando ambos não são apreciados) ou mista (quando um filme do par é apreciado e o outro não).

Os autores observaram que as três camadas tinham características diferentes, que não são observadas quando se considera a projeção global (todas as conexões sem a distinção positiva e negativa). Na projeção global as diferentes avaliações dos usuários são misturadas e, como consequência, essa rede se torna não informativa. No entanto, na perspectiva da rede multiplex, as camadas positivas e negativas apresentam um alto coeficiente de agrupamento, indicando que os usuários tendem a gostar (ou não gostar) de grupos de filmes semelhantes; por outro lado, classificações mistas criam estruturas mais heterogêneas, caracterizadas por comunidades maiores, porém mais esparsas. Outras comunidades online que também foram estudadas utilizando as redes multicamadas, são o *Flickr* [33], o *Facebook* [34] e o *Youtube* [35].

De acordo com Boccaletti e colaboradores [2] os sistemas de transporte também

são candidatos naturais para uma representação de rede multicamadas, uma vez que pode existir vários meios de transporte entre dois locais determinados. Nesse tipo de sistema, as conexões entre vértices diferentes como vôos diretos ou trens, são explícitos, exigindo pouco esforço para a construção da rede. Além disso, os sistemas de transporte são geralmente interligados, sendo a abordagem multicamadas a escolha natural para os estudos de intermodalidade.

Aleta e colaboradores [36] mostraram o impacto que as trocas de transporte podem ter na mobilidade dos indivíduos em uma cidade. Para isso, eles construíram redes multiníveis de nove cidades da Europa de diferentes tamanhos. Nelas, as camadas são os diferentes meios de transporte e os vértices se referem aos pontos de parada do respectivo meio de transporte. As arestas intracamadas dessas redes são estabelecidas quando o itinerário de duas linhas diferentes passam pelos mesmos pontos de parada. Já as arestas intercamadas conectam paradas de diferentes meios de transporte que estão dentro de um raio de 150m. Os autores consideraram redes multiníveis, uma vez que os vértices não existem em todas as camadas.

Além dos exemplos já citados, podemos encontrar aplicações de redes multicamadas em congestionamento e tráfego [37], epidemia [38, 39], dinâmica evolutiva [40], detecção de comunidade [41], mecânica estatística [42], economia [43, 44], ecologia [45] entre outros.

## 1.5 Índices Incidência, Fidelidade, Incidência-Fidelidade para redes de palavras

Nas seções anteriores fizemos um resumo sobre o estudo das redes. Nelas, dissertamos sobre alguns índices gerais utilizados para caracterizar os sistemas descritos em forma de redes. Além disso, apresentamos uma ferramenta muito bem estabelecida na literatura para se trabalhar com redes dinâmicas, os GVT. E por fim, discorreremos sobre as redes compostas por mais de uma camada, as redes multicamadas. Nesta seção, iremos tratar sobre a construção de redes de palavras a partir de discursos orais, sendo que a importância dessa discussão se deve ao fato de utilizarmos índices definidos no estudo desse tipo de rede para o desenvolvimento dos índices de sincronização entre camadas propostos por esse trabalho.

Entre tantas possibilidades de aplicações, as redes têm sido utilizadas também para analisar redes de associações de palavras que emergem de discursos orais [6]. Sendo essas, formadas por palavras, os vértices, e as relações entre elas, as arestas.

Em [6], Teixeira e colaboradores geraram redes de palavras a partir de 12 discursos orais, nos quais cada participante discursou sobre si mesmo durante uma hora. Para a

construção dessas redes, os autores recorreram ao conceito de Incidência<sup>3</sup> proposto por Nelson e colaboradores [46], e também propuseram mais dois índices: a Fidelidade e a Incidência-Fidelidade.

Nessas redes, as palavras são os vértices e as arestas são estabelecidas quando o índice de incidência-fidelidade entre pares de palavras excede um dado limiar. Os autores encontraram que a distribuição de frequência de pares de palavras segue uma lei de potência. Além disso, todas as redes obtidas apresentaram topologias com características de redes de pequeno mundo. Encontraram ainda, que para algumas redes com maior número de vértices, as distribuições de grau apresentam características de uma lei de potência, sugerindo uma topologia de redes livres de escala. A seguir veremos com maiores detalhes os índices utilizados.

### 1.5.1 Incidência

Teixeira e colaboradores [6] definiram o índice Incidência  $I$  como a frequência em que um par de palavras ocorre no decorrer de um discurso. Utilizando a teoria dos conjuntos é possível reescrever matematicamente esse conceito, e para isso considera-se que o conjunto formado por todas as sentenças em que uma dada palavra  $\varphi$  aparece é descrito por  $C_\varphi$ . De modo análogo, o conjunto formado por todas as sentenças em que uma palavra  $\psi$  aparece é descrito por  $C_\psi$ . Define-se que o subconjunto formado por todas as sentenças que ocorre o par de palavras  $\varphi$  e  $\psi$  é dado pela intersecção desses conjuntos,  $C_p \equiv C_\varphi \cap C_\psi$ . O número de elementos de  $C_p$  é dado pela sua cardinalidade, isto é  $S_p \equiv |C_\varphi \cap C_\psi|$ . Dessa forma, pode-se escrever a Incidência como:

$$I_{\varphi,\psi} \equiv \frac{|C_\varphi \cap C_\psi|}{|\cup_{i=1}^{N_p} C_i|}$$

onde  $|\cup_{i=1}^{N_p} C_i| = N_s$  é o número total de sentenças do texto. Assim, temos:

$$I_{\varphi,\psi} \equiv \frac{S_p}{N_s} \quad (1.18)$$

O cálculo da Incidência é realizado para todos os pares de palavras do conjunto de sentenças. A Incidência varia de 0 até 1, quanto mais se aproxima do valor máximo, maior é a frequência do par de palavras ocorrer juntas no discurso.

### 1.5.2 Fidelidade

O índice Fidelidade  $Fi$  calcula a probabilidade de coocorrência de um par de palavras tomando apenas o subconjunto de sentenças onde pelo menos uma das palavras

<sup>3</sup> Originalmente, Nelson e colaboradores denominaram esse conceito como Força. No entanto, para não ser confundido com o conceito de força Newtoniana, Teixeira e colaboradores passaram a usar a denominação Incidência. Neste trabalho também será adotado essa nomenclatura.

investigadas ocorreu durante o discurso. Matematicamente é definido como:

$$Fi_{\varphi,\psi} \equiv \frac{|C_{\varphi} \cap C_{\psi}|}{|C_{\varphi} \cup C_{\psi}|} \quad (1.19)$$

Os conjuntos formados pelas sentenças em que ocorrem as palavras  $\varphi$  e  $\psi$  são não disjuntos, o que significa que podem existir elementos em comum entre os dois conjuntos. Da teoria dos conjuntos sabe-se que é possível escrever a cardinalidade da união dos conjuntos não disjuntos da seguinte forma:

$$|C_{\varphi} \cup C_{\psi}| = |C_{\varphi}| + |C_{\psi}| - |C_{\varphi} \cap C_{\psi}| \quad (1.20)$$

Considerando que  $S_{\varphi} = |C_{\varphi}|$  e  $S_{\psi} = |C_{\psi}|$  escrevemos a Equação 1.20 como:

$$|C_{\varphi} \cup C_{\psi}| = S_{\varphi} + S_{\psi} - S_p \quad (1.21)$$

E assim, a Equação 1.21 toma a seguinte forma:

$$Fi_{\varphi,\psi} \equiv \frac{S_p}{S_{\varphi} + S_{\psi} - S_p} \quad (1.22)$$

Esse índice também está compreendido no intervalo entre 0 e 1.

### 1.5.3 Incidência-Fidelidade

E, por fim, o índice Incidência-Fidelidade  $IFi$  dado pela união dos índices anteriores:

$$IFi_{\varphi,\psi} \equiv I_{\varphi,\psi} \times Fi_{\varphi,\psi}$$

$$IFi_{\varphi,\psi} = \frac{S_p}{N_S} \times \frac{S_p}{S_{\varphi} + S_{\psi} - S_p} = \frac{S_p^2}{N_S(S_{\varphi} + S_{\psi} - S_p)} \quad (1.23)$$

Esse índice representa a probabilidade de ocorrência de um par de palavras no contexto do discurso como um todo e a probabilidade dessas palavras sempre ocorrerem juntas. O índice Incidência-Fidelidade também assume um valor entre 0 e 1.

## 2 Sincronização em redes multicamadas variantes no tempo

### Resumo

Neste trabalho, propomos um método para medir a sincronização intercamadas de sistemas em que as suas conexões mudam com o decorrer do tempo e que são interligados através de diferentes tipos de relações. Na construção deste método, os sistemas são representados por redes multicamadas variantes no tempo, composta por duas camadas sem interconexões entre elas e com número fixo de vértices. O produto desse método é uma rede de Incidência-Fidelidade intercamadas, que indica a intensidade da sincronização das arestas durante um intervalo de tempo. Para isso, consideramos a coocorrência das arestas intracamadas tanto em relação ao tempo de vida do sistema, quanto em relação as frequências de ocorrência em cada camada. Realizamos duas validações: na primeira investigamos a confiabilidade do algoritmo, verificando se as sincronizações intercamadas obtidas computacionalmente para GVT multicamadas aleatórios, estão de acordo com as esperadas matematicamente. E, na segunda, utilizamos o oscilador de Rössler acoplado para gerar séries temporais que representam a presença de duas arestas, uma em cada camada dos GVT multicamadas, e avaliamos a relação entre os valores dos acoplamentos preestabelecidos e os índices de sincronização intercamadas obtidos. Para a aplicação do nosso método, consideramos 17 anos de dados das variáveis climáticas: evapotranspiração, NDVI, precipitação e temperatura, para todos os municípios do estado da Bahia. Para as variáveis individuais evapotranspiração e NDVI, obtivemos redes REAs com módulos super conectados e alto coeficiente de aglomeração na região centro-oeste, que é uma região marcada por sua forte produção agrícola. Para as demais regiões, os municípios fazem poucas conexões ou não se conectam. Para a rede REA de precipitação encontramos que a região mais conectada é a Mata Atlântica, sendo essa região a que apresenta os maiores índices pluviométricos da Bahia. E, por fim, para a rede REA de temperatura obtivemos uma rede em que as conexões são distribuídas de forma homogênea entre os municípios do estado, o que pode ser justificado pelo fato do clima da Bahia apresentar temperaturas elevadas e baixa amplitude térmica. Para as redes de Incidência-Fidelidade geradas a partir dos pares dessas variáveis climáticas, observamos que o padrão de conexão está relacionado com as configurações das redes REAs individuais e com as relações entre as definições dessas variáveis. Para a rede de Incidência-Fidelidade intercamadas do par evapotranspiração-NDVI encontramos uma rede com módulo super conectado na região centro-oeste, assim como nas redes REAs individuais. Esse mesmo padrão também é encontrado na rede de Incidência-Fidelidade intercamadas do par evapotranspiração-



temperatura. Em contrapartida, para os pares de variáveis em que as redes REAs individuais apresentam padrões de conexão em regiões diferentes, as redes de Incidência-Fidelidade intercamadas correspondentes, apresentam conexões distribuídas por todo estado, como é o caso da evapotranspiração-precipitação e NDVI-precipitação. Por fim, as redes de Incidência-Fidelidade intercamadas geradas pelo nosso método, apresentam predominância de ligações de curta distância e também a presença de ligações de longa distância que indicam a dependência não-local das sincronizações intercamadas das redes multicamadas variantes no tempo.

## Palavras-chave:

Redes complexas; sincronização de redes multicamadas; redes variantes no tempo; variáveis climáticas.

## 2.1 Introdução

As redes complexas são objetos matemáticos com a capacidade de codificar informações importantes sobre um sistema utilizando uma estrutura formada por vértices e arestas [47]. As redes podem ser vistas como um dos desenvolvimentos mais prósperos das últimas décadas, uma vez que a sua aplicabilidade compreende uma infinidade de sistemas diferentes [48], que vão desde o estudo de controle de pandemia [49] até a dinâmica do tráfego aéreo, onde os aeroportos e os itinerários dos voos formam uma rede complexa [50].

Os avanços nessa área favoreceram o desenvolvimento de diversas métricas para auxiliar no entendimento e na caracterização dos sistemas descritos como tal. A construção de uma determinada métrica pode atender as particularidades de um sistema específico ou de sistemas de modo geral. Em especial, no estudo de redes de associação de palavras, Teixeira e colaboradores [6] construíram redes monocamadas, onde todas as conexões são centradas em apenas uma rede, baseadas em discursos orais. E para a sua análise, desenvolveram uma métrica denominada Incidência-Fidelidade, que foi definida a partir das probabilidades de duas palavras serem pronunciadas consecutivamente em relação a todo o discurso e em relação as vezes que ao menos uma delas é pronunciada.

Além das métricas, o progresso do estudo das redes trouxe outras contribuições importantes, como o conceito de redes dinâmicas e redes multicamadas. É notório que as conexões entre os componentes de um sistema podem tanto não alterar no decorrer do tempo, quanto apresentar mudanças nos seus padrões de conexão. Quando não há alterações, chamamos essas redes de redes estáticas, do contrário, as redes são denominadas redes dinâmicas.

Existem diferentes metodologias que buscam descrever adequadamente as redes dinâmicas. Holme e Saramäki [16] desenvolveram as redes temporais, que é uma rede onde as arestas indicam os instantes de tempo nos quais as conexões ocorreram. Nos grafos temporais propostos por Kostakos [17], os sistemas são representados por redes em que os seus vértices são distribuídos em uma espécie de "linha do tempo" de acordo com o momento em que realizam conexões. Nesta metodologia, as arestas são direcionadas e podem ser ponderadas e pontilhadas, a depender do tipo de conexão efetivada e do momento em que as conexões são realizadas, respectivamente. Outra metodologia importante é a dos grafos variantes no tempo proposta por Nicosia e Tang [1], onde cada instante de tempo considerado é representado por uma sequência ordenada de grafos estáticos.

Embora os avanços no campo das redes permitam descrever um número cada vez maior de sistemas das mais diversas áreas do conhecimento, nos últimos anos, as limitações presentes na abordagem tradicional de uma rede monocamada, onde diferentes relações podem ser tomadas como equivalentes, tornaram-se cada vez mais evidentes. Bancos de dados com níveis de resolução cada vez mais altos sugerem que os sistemas

do mundo real seriam melhor descritos por uma rede composta por múltiplas camadas (*redes multicamadas*), capaz de representar as distintas interações que compõem um sistema [2, 3, 51, 52].

Em relação as redes multicamadas citadas anteriormente, elas podem ser definidas, de maneira simplista, como uma rede onde cada tipo de relação entre os componentes do sistema é representado por uma camada diferente. Esse tipo de rede oferece uma arquitetura importante para a descrição mais realista de diversos sistemas onde os processos que acontecem em uma rede podem afetar outras redes [53]. Alguns sistemas que podem ser facilmente descritos como uma rede com múltiplas camadas são: as redes sociais, onde os indivíduos são conectados e categorizados por diferentes tipos de relações [54], redes ecológicas, onde cada camada representa um tipo diferente de relação entre os membros de uma mesma espécie [45], redes de transporte, onde cada camada representa um tipo de transporte diferente que conecta os mesmos lugares [55, 56], entre outros sistemas.

Uma das linhas de investigação no estudo das redes multicamadas é entender como acontece a sincronização entre as diferentes camadas dessas redes. Em geral, os processos de sincronização são caracterizados por uma organização espontânea em um estado cooperativo das unidades do sistema [57]. Alguns exemplos de sincronização do cotidiano, são: vaga-lumes piscando em uníssono, atividades neurais ocorrendo no cérebro ou os aplausos de uma plateia. Vários tipos de sincronização já foram estudadas, a sincronização completa [58], sincronização de fase [59] e sincronização de cluster [60], são apenas alguns dos tipos.

O objetivo desse trabalho é desenvolver um método inovador para calcular a sincronização intercadas de sistemas que podem ser avaliados a partir de diferentes camadas e que as conexões entre os componentes de cada camada mudam no decorrer do tempo. Para isso, partimos do conceito do índice Incidência-Fidelidade para redes monocamadas estáticas proposto por Teixeira e colaboradores [6] e propomos um índice de sincronização para redes compostas por duas camadas que mudam com o tempo. Esse novo índice é definido a partir das probabilidades de coocorrência das arestas intracamadas em relação ao tempo de vida do sistema e em relação as frequências dessas arestas nas camadas. O produto desse método é uma rede monocamada ponderada com as respectivas probabilidades das arestas.

Para o desenvolvimento desse método consideramos os conceitos de redes, redes multicamadas, redes variantes no tempo e o índice de caracterização de redes monocamadas propostos por Teixeira e colaboradores [6]. Validamos a confiabilidade do algoritmo implementado utilizando redes em que as camadas foram construídas a partir do modelo de redes aleatórias de Erdős e Rényi  $G(N, p)$  [13] e comparamos os valores das sincronizações obtidos com o algoritmo e os esperados algebricamente. Realizamos ainda, a validação do método utilizando séries temporais obtidas com o oscilador de Rössler acoplado [61] e comparamos os resultados obtidos com o esperado para diferentes valores do acoplamento.

Aplicamos o método em dados de variáveis climáticas (evapotranspiração, índice de vegetação por diferença normalizada, precipitação e temperatura) das 417 cidades do estado da Bahia, no decorrer de 17 anos. E, então, analisamos tanto as redes individuais de cada variável climática, quanto as redes de sincronização intercamadas entre pares específicos dessas variáveis e as analisamos a partir dos biomas de cada região.

Nas próximas seções descrevemos o método proposto, juntamente com as suas validações e aplicação. Finalizamos o capítulo com as conclusões sobre o trabalho, seguida dos apêndices relativos a aplicação.

## 2.2 O Método

O método de sincronização de redes multicamadas variantes no tempo que estamos propondo considera os seguintes pontos fundamentais: a coocorrência das arestas nas diferentes camadas durante todo o tempo de vida dos GVT e, a coocorrência das arestas ao considerar a assiduidade delas em pelo menos uma das camadas.

Para elaborar a formulação matemática necessária para o desenvolvimento deste trabalho, partimos das definições de grafos variantes no tempo [1] e de redes multicamadas [2–4]. Assim, definimos uma rede multicamadas  $\mathcal{M} = (g, c)$  onde  $g$  é o conjunto das camadas, que por sua vez são compostas de grafos variantes no tempo e  $c$  é o conjunto das arestas intercadas.

O conjunto das camadas é dado por  $g = \{\mathcal{G}_\alpha \mid \alpha \in \{1, 2, \dots, P\}\}$ , onde  $\mathcal{G}_\alpha$  é o grafo variante no tempo referente a camada  $\alpha$  e  $P$  indica o número máximo de camadas. Para o nosso propósito, escrevemos um grafo variante no tempo como uma sequência ordenada  $\mathcal{G}_\alpha = \{G_\alpha(t_1), G_\alpha(t_2), \dots, G_\alpha(t_\Gamma)\}$  composta por  $\Gamma$  grafos, que indica o tempo de vida do sistema, com  $N$  vértices fixos. Nesta notação, o grafo  $G_\alpha(t_s)$  refere-se ao estado da camada  $\alpha$ , no tempo  $t_s$ , para  $s = 1, 2, \dots, \Gamma$ .

Um grafo  $G_\alpha(t_s) = (V_\alpha, E_\alpha(t_s))$  possui um conjunto de vértices  $V_\alpha = \{v_1^\alpha, v_2^\alpha, \dots, v_{N_\alpha}^\alpha\}$  e um conjunto particular de arestas,  $E_\alpha(t_s)$ , para cada  $t_s$ . A matriz de adjacência dependente do tempo  $A^{[\alpha]}(t_s) = a_{i,j}^\alpha(t_s) \in \mathbb{R}^{N_\alpha \times N_\alpha}$ , para esse grafo é dada por:

$$a_{i,j}^\alpha(t_s) = \begin{cases} 1, & \text{se } v_i^\alpha \text{ e } v_j^\alpha \text{ estão conectados} \\ 0, & \text{caso contrário} \end{cases} \quad (2.1)$$

O conjunto das arestas intercadas é dado por  $c = \{E_{\alpha\beta}(t_s) \subseteq V_\alpha \times V_\beta; \alpha \text{ e } \beta \in \{1, 2, \dots, P\} \text{ com } \alpha \neq \beta \text{ e } s \in \{1, 2, \dots, \Gamma\}\}$ .

*Inicialização:* No desenvolvimento desse método consideramos uma rede multicamadas  $\mathcal{M} = (g, c)$ , onde o conjunto das camadas é dado por  $g = \{G, F\}$ , em que  $G = \mathcal{G}_1$  e  $F = \mathcal{G}_2$ . Os grafos variantes no tempo dessas camadas possuem o mesmo tempo de vida  $\Gamma$  e mesmo número de vértices. Nesse método não existem arestas entre as camadas, dessa forma o conjunto  $c$  é vazio.

A Figura 10 traz um exemplo de uma rede multicamadas variante no tempo com as características consideradas no método. Nesta figura, temos uma rede com 2 camadas e tempo de vida igual a 5, na qual não existem arestas intercadas.

*Calculando os índices de sincronização intercadas:* Definimos três índices de sincronização das arestas, denominados Incidência ( $I_{i,j}$ ), Fidelidade ( $F_{i,j}$ ) e Incidência-Fidelidade ( $IF_{i,j}$ ).

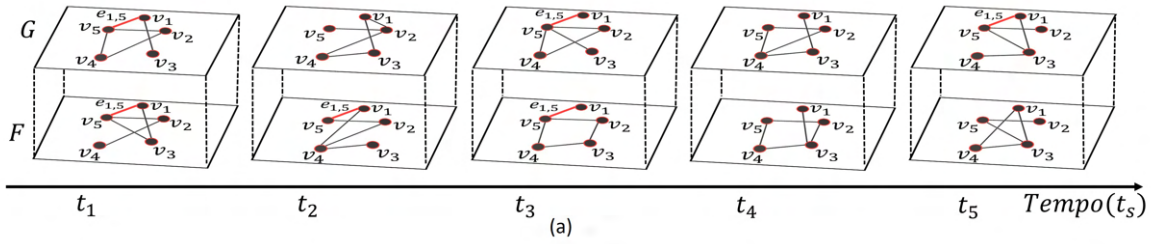


Figura 10 – Rede multicamadas variante no tempo com  $P = 2$  e  $\Gamma = 5$ .

*Incidência:* É definida como uma forma de medir quantas vezes uma determinada aresta  $e_{i,j}$  apareceu nas camadas  $G$  e  $F$  no instante de tempo  $t_s$  ao longo do tempo de vida  $\Gamma$  dos GVT.

Matematicamente:

$$I_{i,j} = \frac{\sum_{s=1}^{\Gamma} a_{i,j}^G(t_s) a_{i,j}^F(t_s)}{\Gamma} \quad (2.2)$$

onde,

$$a_{i,j}^x(t_s) = \begin{cases} 1, & \text{se existir uma aresta } e_{i,j} \text{ no tempo } t_s \text{ na camada } x \\ 0, & \text{caso contrário} \end{cases} \quad (2.3)$$

Após calcular a Incidência para todas as arestas, construímos uma rede ponderada com os valores da Incidência de cada aresta. Nela, os pesos estão compreendidos no intervalo de 0 até 1 e indicam o quanto as arestas coocorreram nas camadas  $G$  e  $F$  no mesmo instante de tempo  $t_s$  durante todo o GVT. O peso 0, indica que as arestas não coocorrem em nenhum momento dos GVT, já o peso máximo indica que elas coocorrem durante todo o tempo de vida dos GVT o que implica em uma sincronização intercadas perfeita.

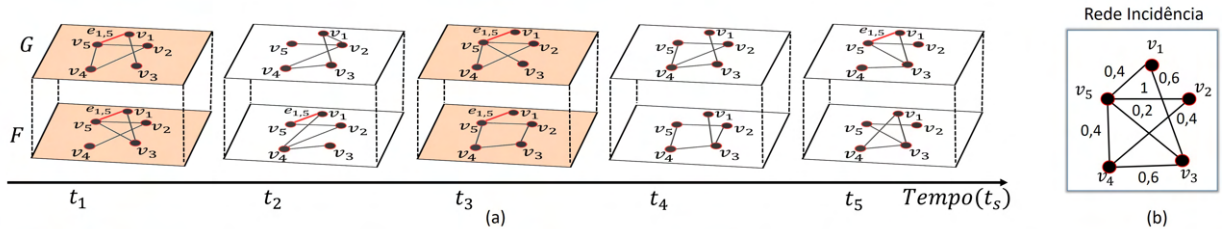


Figura 11 – (a) Rede multicamadas com 5 vértices em cada camada e tempo de vida dos GVT igual a 4. (b) Rede de Incidência para os GVT considerados.

A Figura 11(a) apresenta um exemplo de uma rede multicamadas com 5 vértices em cada camada e tempo de vida dos GVT igual a 5 e a Figura 11(b) traz a sua respectiva rede de Incidência. Nesta rede temos como destaque a aresta  $e_{1,5}$  com valor da Incidência igual a 0,4, indicando assim que essa aresta coocorreu nas camadas  $G$  e  $F$  em menos da metade

do tempo de vida dos GVT. Já a aresta  $e_{2,5}$  possui o valor máximo da Incidência, ou seja, essa aresta coocorreu nas camadas  $G$  e  $F$  durante todo tempo considerado, como pode ser verificado na figura. Por outro lado, a aresta  $e_{1,2}$  não aparece na rede de Incidência, embora ela tenha acontecido no instante  $t_2$  da camada  $G$ , ela não aconteceu nas duas camadas no mesmo instante de tempo.

*Fidelidade:* O cálculo desse índice considera o número de vezes que uma determinada aresta  $e_{i,j}$  apareceu, e dessas, quantas ela apareceu ao mesmo tempo  $t_s$  nas camadas  $G$  e  $F$ . Matematicamente, podemos escrever:

$$F_{i,j} = \frac{\sum_{s=1}^{\Gamma} a_{i,j}^G(t_s) a_{i,j}^F(t_s)}{\sum_{s=1}^{\Gamma} a_{i,j}^G(t_s) + \sum_{s=1}^{\Gamma} a_{i,j}^F(t_s) - \sum_{s=1}^{\Gamma} a_{i,j}^G(t_s) a_{i,j}^F(t_s)} \quad (2.4)$$

Os valores de  $G$  e  $F$  novamente são definidos de acordo com a condição (2.3). A Fidelidade também está compreendida entre os valores de 0 a 1. Após calcular esse índice para todas as arestas, podemos escrever a rede de Fidelidade. O índice Fidelidade considera apenas as frequências individuais e coocorrência das arestas nas camadas.

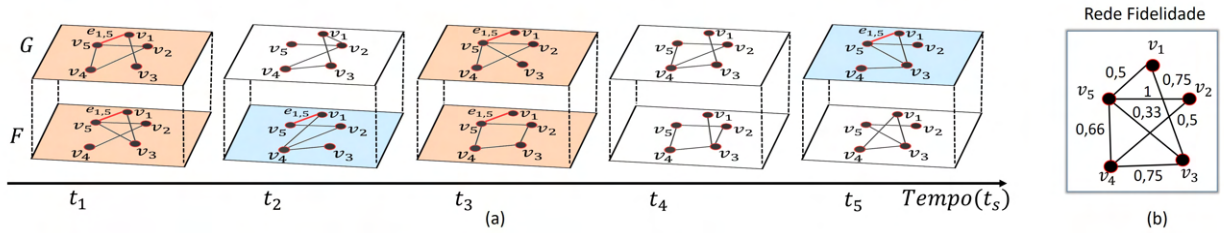


Figura 12 – (a) Rede multicamadas com 5 vértices em cada camada e tempo de vida dos GVT igual a 5. (b) Rede de Fidelidade para os GVT considerados.

A Figura 12(a) apresenta a mesma rede multicamadas variante no tempo do exemplo anterior, no entanto agora é calculada a rede de Fidelidade desses GVT, Figura 12(b). A aresta  $e_{1,5}$  possui o valor da Fidelidade igual a 0,5, como este índice considera somente os instantes em que as arestas apareceram, ainda que individualmente, o tempo  $t_4$  é desconsiderado nos cálculos da Equação (2.4). Assim o valor da Fidelidade dessa aresta é maior do que o valor da Incidência, porque o espaço amostral nesse caso é menor.

Uma situação importante que pode gerar uma interpretação equivocada da sincronização intercadas, é quando temos uma aresta que aparece pouquíssimas vezes nos GVT, mas todas as vezes que aparece, ela coocorre nas duas camadas. Essa situação gera valores altos para a Fidelidade, mesmo que essa aresta tenha aparecido apenas uma vez nos GVT. Devido a situações como essa, introduzimos o índice de Incidência-Fidelidade garantindo que o valor máximo da sincronização intercadas, ou seja 1, expresse realmente a sincronização máxima dos GVT.

*Incidência-Fidelidade:* É o produto dos índices Incidência e Fidelidade e como já mencionado, visa eliminar qualquer interpretação errada dos valores da sincronização analisados individualmente. Dessa forma, o índice Incidência-Fidelidade expressa a sincronização intercadas das arestas tanto em relação ao tamanho dos GVT como das frequências individuais. Matematicamente escrevemos o índice Incidência-Fidelidade, como

$$IF_{i,j} = I_{i,j} \times F_{i,j} \quad (2.5)$$

Assim como para os outros índices, também construímos uma rede de Incidência-Fidelidade. As análises discutidas no decorrer desse trabalho são baseadas nesse índice especificamente.

Rede de Incidência-Fidelidade

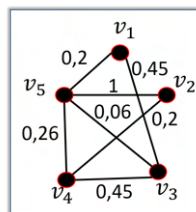


Figura 13 – Rede de Incidência-Fidelidade para a rede multicamadas dos exemplos anteriores.

A Figura 13 apresenta a rede de Incidência-Fidelidade para a rede multicamadas consideradas anteriormente.

Na construção desse método, admitimos que não há atraso entre as camadas. Em desenvolvimentos futuros podemos incluir o atraso como uma possibilidade a ser considerada.

O método proposto foi implementado na linguagem de programação C++ e foi utilizado o IDE multi-plataforma Qt Creator 4.13.0.

## 2.3 Validações

Nesta seção propomos duas formas de validação do método que desenvolvemos. Primeiro, validamos o código, para isso aplicamos o método em GVT compostos por redes multicamadas construídas a partir de um processo aleatório e comparamos os resultados obtidos matematicamente com os resultados obtidos por meio de simulações computacionais. E por fim, validamos o modelo utilizando séries temporais com níveis de acoplamento previamente estabelecidos, analisamos o comportamento dos índices de sincronização intercadas.



### 2.3.1 Sincronização em Redes Multicamadas Aleatórias Variantes no Tempo

Nesta etapa analisamos a sincronização intercadas das arestas em redes multicamadas onde as intraconexões foram realizadas a partir de um processo aleatório, levando em conta que as conexões entre os vértices de cada camada mudam no decorrer do tempo.

Para isso, para cada instante de tempo  $t_s$  dos GVT, consideramos uma rede composta por duas *camadas aleatórias* que foram construídas a partir do modelo de redes aleatórias de Erdős e Rényi  $G(N, p)$  [13] e calculamos a sincronização entre essas camadas.

Na construção das camadas com conexões aleatórias definimos o número de vértices  $N$  e a probabilidade de conexão  $p$  entre os vértices. De acordo com o modelo de redes aleatórias de Erdős e Rényi [13], o número  $m$  de conexões que serão realizadas em cada camada pode ser calculado por meio da expressão:

$$m = pm_{max} \quad \text{sendo} \quad m_{max} = \frac{N(N-1)}{2} \quad (2.6)$$

onde  $m_{max}$  é o número máximo possível de arestas pra uma camada com  $N$  vértices.

Após a construção da rede multicamadas aleatória variante no tempo, pode-se calcular os valores dos índices de Incidência, Fidelidade e Incidência-Fidelidade de acordo com o nosso método descrito na subseção 2.2. Neste caso específico em que as conexões são realizadas por meio de um processo aleatório é possível encontrar as expressões analíticas para estes índices, que podem então ser comparados com os valores obtidos pelas simulações numéricas.

De acordo com a definição da Incidência, Equação 2.2, podemos concluir que se a aresta  $e_{i,j}$  da camada  $G$  tem probabilidade  $p$  de ocorrer e a aresta  $e_{i,j}$  da camada  $F$  tem probabilidade  $q$  de ocorrer, então ao somarmos o produto das camadas  $a_{i,j}^G(t_s)$  e  $a_{i,j}^F(t_s)$  para todo o intervalo de tempo  $\Gamma$  e dividirmos por  $\Gamma$  temos a probabilidade de coocorrência das arestas ao longo do tempo que será equivalente ao produto das probabilidades de formação dos grafos aleatórios  $pq$ .

$$\begin{aligned} I_{i,j} &= \frac{\sum_{s=1}^{\Gamma} a_{i,j}^G(t_s) a_{i,j}^F(t_s)}{\Gamma} \\ &= pq \end{aligned}$$

De forma análoga, o índice Fidelidade fica:

$$\begin{aligned} F_{i,j} &= \frac{\sum_{s=1}^{\Gamma} a_{i,j}^G(t_s) a_{i,j}^F(t_s)}{\sum_{s=1}^{\Gamma} a_{i,j}^G(t_s) + \sum_{s=1}^{\Gamma} a_{i,j}^F(t_s) - \sum_{s=1}^{\Gamma} a_{i,j}^G(t_s) a_{i,j}^F(t_s)} \\ &= \frac{pq}{p + q - pq} \end{aligned}$$

E por fim, a Incidência-Fidelidade:

$$\begin{aligned} IF_{i,j} &= I_{i,j} \times F_{i,j} \\ &= pq \frac{pq}{p+q-pq} \\ &= \frac{p^2q^2}{p+q-pq} \end{aligned}$$

Nas expressões acima, consideramos  $p$  e  $q$  diferentes de zero, uma vez que para esse valor as camadas não possuem arestas.

O gráfico da Figura 14 tem o propósito de demonstrar que os resultados esperados nas simulações estão de acordo com os obtidos matematicamente.

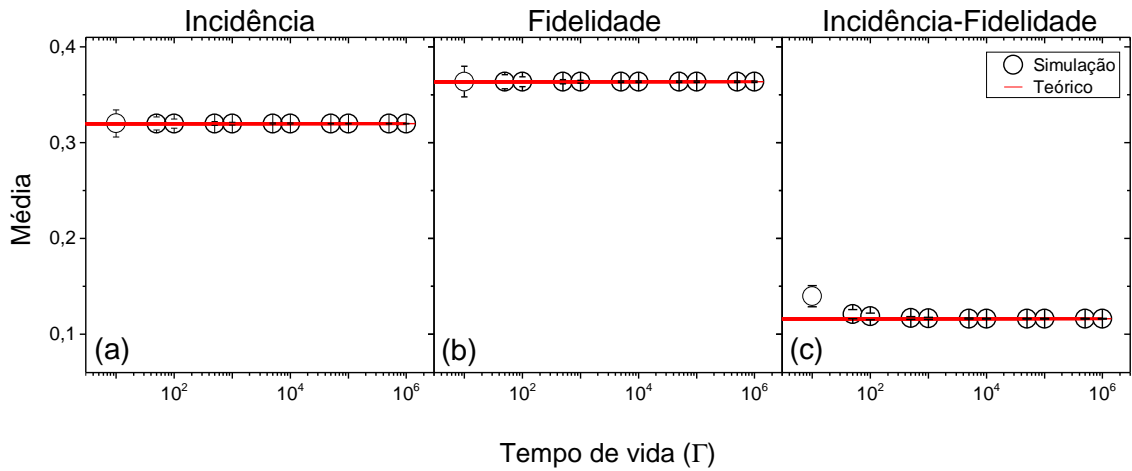


Figura 14 – Simulações para redes com duas camadas e diferentes tempos de vida dos GVT ao considerar um total de 100 amostras. Na construção dos GVT aleatórios a probabilidade da camada  $G$  é igual a 0,4 e da camada  $F$  é igual a 0,8. Valor médio dos índices (a) Incidência, (b) Fidelidade e (c) Incidência-Fidelidade.

Nesta etapa consideramos um total de 100 amostras para diferentes tempos de vida dos GVT, com  $\Gamma$  variando de 10 até  $10^6$ . A rede multicamada de cada amostra é composta por 50 vértices em cada camada. Estabelecemos que as probabilidades de conexões na camada  $G$  é igual a 0,4 e na camada  $F$  é igual a 0,8, o que implica em 20 e 40 intraconexões, respectivamente. Para cada valor de  $\Gamma$  construímos as redes com as médias dos índices de sincronização sobre o número de amostras considerados. E, a partir dessa rede, calculamos a média e o desvio padrão de cada índice em relação ao número de arestas das redes, Figura 14.

Na escolha dos parâmetros: tamanho da amostra, número de vértices e as probabilidades de conexão, consideramos os limites dos recursos computacionais disponíveis.

No entanto, nos certificamos de que essas escolhas fossem suficientes para garantir a estabilidade dos resultados sem custo computacional alto. Utilizamos esse mesmo critério na escolha dos parâmetros referentes a validação com o Oscilador de Rössler acoplado.

Observamos que mesmo existindo flutuações entre as médias dos índices de sincronização intercadas para os diferentes tempos de vida dos GVT, os valores médios da Incidência e da Fidelidade estão dentro do esperado teoricamente, como podemos ver nos gráficos das Figuras 14(a) e (b). Para a Incidência-Fidelidade, quando consideramos o tempo de vida dos GVT relativamente pequeno, a sua média se afasta do valor esperado teoricamente. Nesses casos, o tempo considerado permite maiores flutuações nos valores das médias e o produto dessas flutuações amplifica o valor da média da Incidência-Fidelidade que, por sua vez, se afasta do valor esperado teoricamente, Figura 14 (c).

Em síntese, podemos afirmar que de acordo com o gráfico da Figura 14 e as discussões apresentadas, os valores encontrados nas simulações para os índices de Incidência, Fidelidade e Incidência-Fidelidade considerando redes multicamadas variantes no tempo, nas quais as intraconexões aconteceram de forma aleatória, tendem aos valores preditos analiticamente para diferentes tempos de vida dos GVT. É importante observar ainda que, como os índices de sincronização intercadas não dependem do número de vértices das camadas, podemos encontrar esses mesmos valores ao considerar camadas com as mesmas configurações, mas com número de vértices diferente.

Até este momento, avaliamos a sincronização intercadas a partir das intraconexões que ocorreram de forma aleatória, ou seja, para modelos nulos. Na próxima seção admitimos intraconexões que podemos controlar a sua presença nas diferentes camadas da rede.

### 2.3.2 Oscilador de Rössler Acoplado

Nessa subseção vamos verificar se as redes de sincronização intercadas obtidas com o nosso método estão de acordo com os dados utilizados para construir as redes multicamadas variantes no tempo. Para realizar essa verificação, consideramos que as camadas possuem apenas dois vértices e que existe uma série temporal associada a cada uma delas. Essas séries temporais indicam em quais instantes de tempo do GVT as arestas estão ativas.

Para gerar as séries temporais consideramos o oscilador de Rössler acoplado [61]. Esse oscilador descreve o tipo de sistema que estamos interessados em investigar e a sua implementação é razoavelmente simples. O oscilador de Rössler é um sistema formado por dois elementos em que é possível controlar o nível de acoplamento entre eles. Em princípio, a intensidade do acoplamento reflete nos valores da sincronização entre esses elementos. O nosso objetivo com o oscilador de Rössler é constatar esse comportamento.

O oscilador de Rössler foi introduzido na literatura em 1976 e trata-se de uma simplificação do modelo de Lorenz. Esse oscilador considera um acoplamento na componente  $x$ , e suas equações são da seguinte forma:

$$\begin{aligned}\dot{x}_{1,2} &= -\omega_{1,2}y_{1,2} - z_{1,2} + \varepsilon_1(x_{2,1} - x_{1,2}) \\ \dot{y}_{1,2} &= \omega_{1,2}x_{1,2} + 0,165y_{1,2} \\ \dot{z}_{1,2} &= 0,2 + z_{1,2}(x_{1,2} - 10)\end{aligned}\quad (2.7)$$

onde  $\varepsilon_{2,1}$  são as forças de acoplamento e  $\omega_{1,2}$  são as frequências naturais dos osciladores.

Para resolver as equações de movimento 2.7 utilizamos o método de integração numérica Runge-Kutta de quarta ordem com o tamanho do passo  $dt = 0,1$  e posição inicial aleatória dentro de uma esfera de raio  $r$ . Com essa integração obtivemos duas séries temporais, cada uma delas, representa uma aresta numa dada camada  $n$ . Para calcular a sincronização entre essas arestas, transformamos as suas séries temporais em séries binárias que indicam a presença das arestas para cada instante de tempo  $t'$  do tempo total  $t$ .

Construímos uma série binária obedecendo os seguintes passos: primeiro, escrevemos a série temporal de uma aresta da camada  $n$ , como  $\{x_{t'}^n\}_{t'=1}^t = \{x_1^n, x_2^n, \dots, x_{t-1}^n, x_t^n\}$ , onde  $x_{t'}^n$  é a posição  $x$  do oscilador no tempo  $t'$ . E, então, definimos a média temporal dessa série:

$$X_m^n = \frac{\sum_{t'=1}^t x_{t'}^n}{t}. \quad (2.8)$$

O valor de  $X_m^n$  é utilizado como um parâmetro de referência que nos indica em quais instantes, no decorrer de  $t$ , a aresta está ativa. Essa classificação é realizada com o auxílio de um valor binário,  $e_{t'}^n$ , definido com a seguinte condição:

$$e_{t'}^n = \begin{cases} 1, & \text{se } x_{t'}^n \geq X_m^n \\ 0, & \text{se } x_{t'}^n < X_m^n \end{cases} \quad (2.9)$$

Se a posição  $x_{t'}^n$  do oscilador no tempo  $t'$  for maior ou igual a média temporal da série, a constante  $e_{t'}^n$  recebe o valor 1, indicando a presença da aresta naquele instante de tempo. Caso contrário, é atribuído o valor 0 para  $e_{t'}^n$  afirmando que não há aresta no tempo  $t'$ .

Por fim, para uma dada aresta  $a_{i,j}$  de uma camada  $n$ , construída a partir do oscilador de Rössler acoplado, a série binária que indica a presença dessa aresta no decorrer do tempo  $t$ , é dada por:

$$a_{i,j}^n = \{e_{t'}^n\}_{t'=1}^t \quad (2.10)$$

Em síntese, o oscilador de Rössler acoplado gera duas séries temporais, cada série descreve uma aresta em uma dada camada, que denominaremos  $G$  e  $F$ . Após a conversão dessas séries em séries binárias, assumimos que o tempo  $t'$  das séries binárias é análogo ao tempo  $t_s$  dos grafos variantes no tempo. Por fim, calculamos a sincronização intercadas utilizando os índices definidos nas expressões de 2.2, 2.4 e 2.5.

Em vista disso, simulamos um total de  $a$  osciladores e calculamos a média dos índices de sincronização intercadas para essas amostras, dada como:

$$\bar{S} = \frac{\sum_{a'=1}^a S_{a'}}{a} \quad (2.11)$$

onde  $\bar{S}$  representa a média de um dos índices de sincronização intercadas mencionados acima, e  $S_{a'}$  refere-se ao valor encontrado para o índice de sincronização intercadas  $S$  ao considerar as amostras  $a$ . Calculamos também o desvio padrão para os valores de sincronização intercadas, utilizando a expressão:

$$\sigma = \sqrt{\frac{\sum_{a'=1}^a (S_{a'} - \bar{S})^2}{n - 1}} \quad (2.12)$$

Os gráficos da Figura 15 exibem a média e o desvio padrão dos índices de sincronização ao considerar 500 amostras de pares de séries temporais com diferentes níveis de acoplamento entre elas. Cada série temporal representa a função presença de uma aresta em uma camada, sendo cada camada formada por apenas dois vértices. As séries temporais simuladas possuem 1000 pontos, com posições iniciais aleatórias dentro de uma esfera com um raio igual a 0,5 centrado em  $(0, 0, 0)$ . As frequências naturais dos osciladores recebem os valores  $\omega_1 = 0,81$  e  $\omega_2 = 0,8$  e a força de acoplamento  $\varepsilon_1$  assume valores entre 0 e 1 com passo de crescimento igual a 0,01.

Como podemos observar nos gráficos, a sincronização intercadas cresce à medida que os níveis de acoplamento entre as séries se tornam cada vez maiores. Assim, quanto maior o acoplamento, mais parecidas são as séries e menor é o desvio padrão das médias de sincronização intercadas. Consequentemente, quanto menor o acoplamento, menos parecidas são as séries e menor é a sincronização intercadas entre elas.

No extremo em que o acoplamento entre as séries temporais é zero, as arestas aparecem aleatoriamente nos GVT, em ambas camadas, e os índices de sincronização intercadas apontam valores baixos com desvio na mesma ordem de grandeza das médias obtidas.

Quando o acoplamento é máximo as arestas coocorrem com maior frequência e os índices de sincronização intercadas atingem valores mais altos do que os obtidos

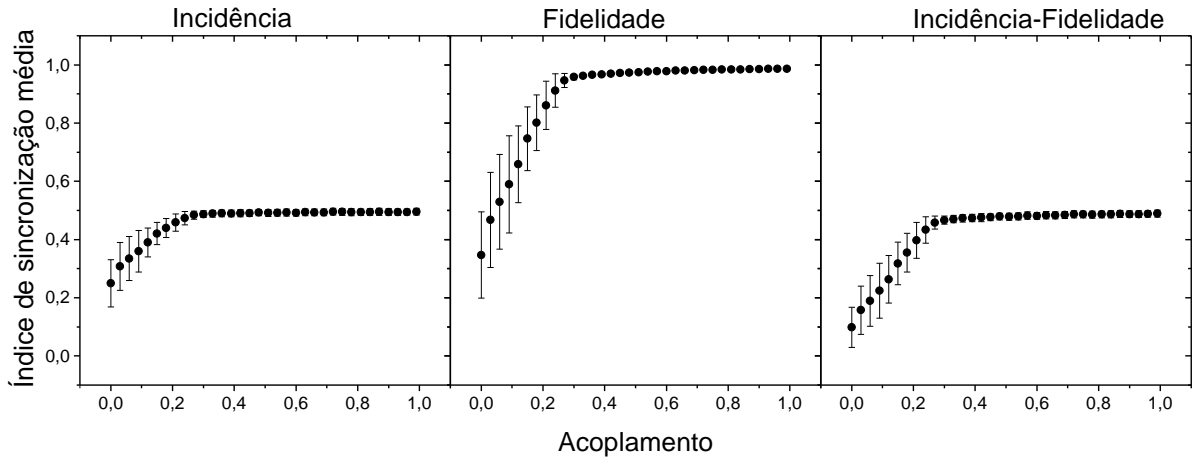


Figura 15 – Índices de sincronização intercamadas para diferentes valores do acoplamento

anteriormente. De acordo com o gráfico da Figura 15, essa sincronização, no entanto, alcança apenas metade dos valores máximos para os índices de Incidência (0, 5) e, por consequência, de Incidência-Fidelidade. A explicação para tal comportamento está na escolha da média como limiar para existência de arestas 2.9. Por definição da média, metade dos valores estarão acima e metade abaixo dela. Assim é esperado que as arestas ocorram em, no máximo, o tempo total  $t$  da série, o que leva a um valor máximo de incidência de 0, 5.

Outra observação importante diz respeito a sincronização intercamadas alcançar os valores máximos ao considerar a força de acoplamento igual ou maior do que 0, 3. Esse comportamento é uma característica do Oscilador de Rössler acoplado. A partir desse valor, ainda que existam pequenas flutuações entre as séries temporais, elas não são detectadas, uma vez que estamos usando valores médios como fator determinante da presença das arestas.

Com isso, podemos afirmar que o método que desenvolvemos está coerente com os resultados esperados.

## 2.4 Aplicação

Nesta seção apresentamos a aplicação do método de Sincronização de Redes Multicamadas Variantes no Tempo em dados de variáveis climáticas. A estrutura dessa seção está dividida em duas subseções. Na primeira subseção descrevemos as variáveis climáticas utilizadas, juntamente com a plataforma que usamos para coletar esses dados e definimos, ainda, os valores dos parâmetros do método de sincronização por motifs utilizado para construir as redes das variáveis climáticas. E, na segunda subseção, apresentamos as redes obtidas para as variáveis climáticas individuais e as redes de sincronização intercamadas obtidas com o nosso método.

### 2.4.1 Materiais e Métodos

Aplicamos o método de Sincronização de Redes Multicamadas Variantes no Tempo em séries temporais de evapotranspiração, índice de vegetação por diferença normalizada - NDVI, precipitação e temperatura para todo o estado da Bahia, em um intervalo de tempo de 17 anos (entre 2001 e 2018).

A Bahia foi escolhida devido a grande diversidade entre os seus municípios, tanto em relação ao clima e a topografia, quanto em relação ao uso e cobertura do solo. Além disso, de acordo com o Instituto Brasileiro de Geografia e Estatística - IBGE [62] a Bahia é o quinto maior estado brasileiro em extensão territorial com cerca de  $564.760 \text{ km}^2$ , e o quarto maior estado com o maior número de municípios, 417. A Bahia possui uma enorme diversidade de vida animal e formações vegetais, desde florestas a ambientes campestres [63,64], sendo composto por três biomas: Caatinga, Cerrado e Mata Atlântica, como pode ser visto na Figura 16.

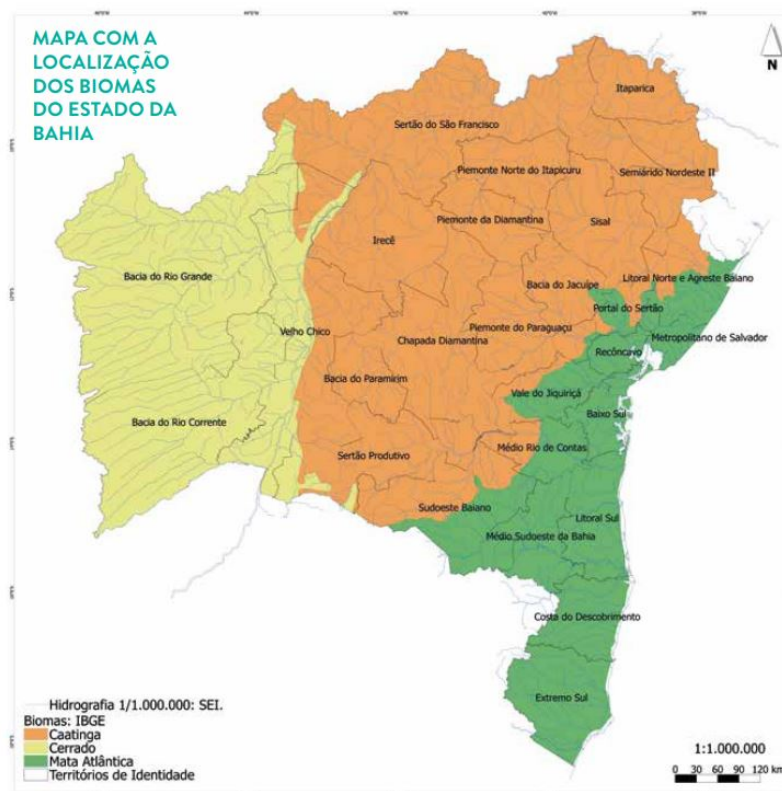


Figura 16 – Mapa da Bahia com seus respectivos biomas sendo representados por diferentes cores: em amarelo claro, o Cerrado; em laranja a Caatinga e em verde a Mata Atlântica. Fonte: [63].

Cada um desses biomas apresenta aspectos distintos, tanto da perspectiva ecológica quanto da sociocultural. No Apêndice B apresentamos algumas características gerais destes biomas com o objetivo de uma melhor análise e compreensão dos resultados obtidos.

Em relação as variáveis climáticas utilizadas nessa aplicação, o critério de escolha

adotado considera tanto a relação que existe entre elas, quanto a importância dessas variáveis na caracterização dos diferentes padrões climáticos dos municípios.

A evapotranspiração é definida como a combinação de dois processos distintos, a evaporação da água diretamente da superfície do solo e a transpiração através dos estômatos das plantas [65]. Além disso, a evapotranspiração afeta a umidade atmosférica, o balanço hídrico de microbacias e o desenvolvimento e rendimento de agriculturas [66]. Um dos fatores que mais afetam a taxa de evapotranspiração é a temperatura, também considerada nesse trabalho, uma vez que, a radiação solar é a maior fonte de energia que possibilita a passagem da fase líquida para gasosa da água [66, 67].

Proposto por Rouse e colaboradores [68], o NDVI trata-se de uma ferramenta rápida e eficiente de detecção de variações na vegetação [69]. Ele permite identificar tanto a presença de vegetação verde e caracterizar sua distribuição espacial, como também a evolução no decorrer do tempo. A grosso modo, o NDVI está relacionado com a fotossíntese das plantas e é afetado, entre outros fatores, pela temperatura e também pela quantidade de água disponível no sistema.

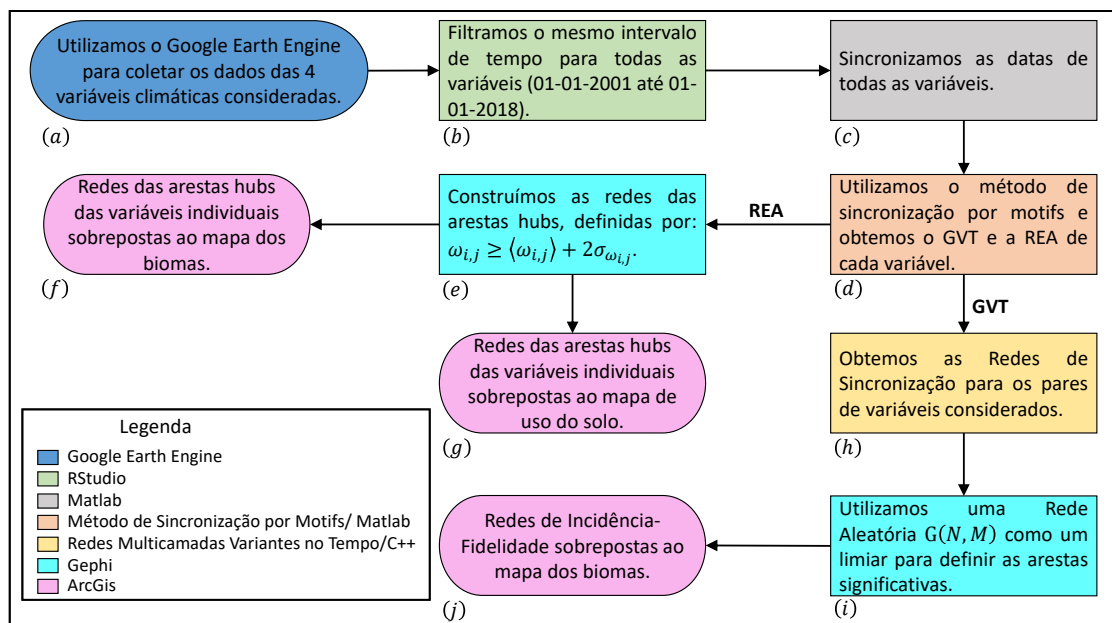


Figura 17 – Fluxograma detalhando as etapas seguidas na aplicação do nosso método com os dados de variáveis climáticas e as ferramentas utilizadas.

As séries temporais referentes a essas variáveis climáticas foram obtidas através de sensoriamento remoto. Na Figura 17 temos um fluxograma descrevendo os passos dessa aplicação, desde a obtenção dos dados até a construção das redes de sincronização intercadas.

A plataforma utilizada na obtenção dos dados foi o *Google Earth Engine-GEE*, como indica a Figura 17(a). O GEE é uma plataforma de processamento de dados geográficos com computação em nuvem. Essa plataforma dispõe de alta capacidade de armazenamento



e processamento de imagens de satélites e permite ao usuário realizar diversas análises espaço-temporais, a nível regional e global, com velocidade muito maior do que às técnicas convencionais [70, 71].

O GEE fornece acesso direto a um catálogo de vários petabytes de imagens de satélite e conjuntos de dados geoespaciais, incluindo todo o catálogo Landsat, MODIS, CHIRPS, bem como arquivos completos de dados de elevação, cobertura da terra, temperatura da superfície do mar e dados climáticos [70, 71].

Para o presente trabalho, criamos rotinas automatizadas no GEE para extrair as séries temporais dos dados das variáveis climáticas levando em consideração a delimitação da poligonal de cada município e, então, tiramos a média dos valores dos pixels das séries temporais que estão no interior de cada uma dessas geometrias.

As séries temporais obtidas passaram por um processo de filtragem para delimitar o intervalo de tempo em comum para todas as variáveis, especificamente, a partir de janeiro de 2001 até janeiro de 2018. O script utilizado foi implementado no RStudio, como descrito na Figura 17(b).

Foi desenvolvido ainda, um script no Matlab para uniformizar as frequências das coletas, ver Figura 17(c). Nele, utilizamos as datas das coletas de NDVI como referência para sincronizar as datas das demais variáveis climáticas. Como os valores dos índices de NDVI variam lentamente, consideramos fazer coletas quinzenais, como mostrado na Tabela 3. Com esse intervalo de tempo é possível observar variações nesse índice.

Variável climática	Frequência dos dados	Operação
Evapotranspiração	semanal	média
NDVI	quinzenal	—
Precipitação	diária	soma
Temperatura	semanal	média

Tabela 3 – Operações realizadas nos dados para transformá-los em medidas quinzenais.

Para transformar os dados originais de precipitação em dados quinzenais, somamos os seus valores diários considerando o intervalo de tempo definido nos dados de NDVI. Já para os dados semanais de evapotranspiração e temperatura, nós consideramos a média das medidas no intervalo de tempo considerado. As operações realizadas para cada variável climática estão apresentadas na Tabela 3.

No final desse processo, temos 4 séries temporais para cada município com datas simultâneas as dos dados de NDVI. A Figura 18 traz um exemplo ilustrativo de fragmentos das séries temporais obtidas para a cidade de Salvador para o ano de 2017.

Em 76, 5% dos 17 anos considerados, verificamos que, para as séries de NDVI, foi realizada apenas uma coleta no mês de outubro. Assim, idealmente, cada série temporal do NDVI deveria ter 408 pontos, no entanto, devido à essa perda, ao fim cada série temporal

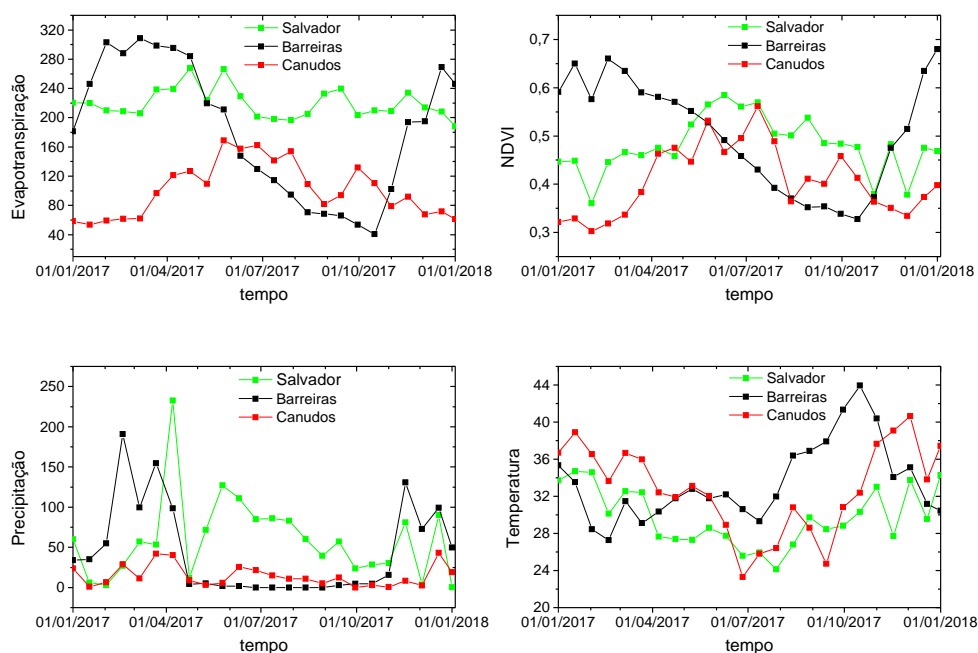


Figura 18 – Séries temporais das variáveis climáticas do ano de 2017. Foram consideradas as seguintes cidades: Salvador, pertencente ao bioma Mata Atlântica (em verde); Barreiras, pertencente ao Cerrado (em preto) e Canudos, cujo bioma é a Caatinga (em vermelho).

foi composta por 392 pontos. Como os dados de NDVI servem como referência, as demais séries temporais também consideram essa quantidade de pontos.

Outros detalhes importantes sobre como as coletas foram realizadas estão especificados abaixo:

- **Evapotranspiração:** o banco de dados de evapotranspiração para o intervalo de tempo considerado foi processado usando a coleção de imagens MOD16A2 V6, com resolução de pixel de 500m e 8 dias [72, 73].
- **Índice de Vegetação por Diferença Normalizada:** determinado através da leitura da reflectância nas bandas do infravermelho próximo e vermelho. O banco de dados NDVI foi processado usando o produto de coleta de imagens MOD13Q1 V6 que fornece um valor de índice de vegetação por pixel, onde cada pixel tem resolução espacial de 250m e resolução temporal de 16 dias [74, 75].
- **Precipitação:** o banco de dados de precipitação foi processado usando a coleção de imagens CHIRPS( Precipitação infravermelha do grupo de riscos climáticos com dados da estação). O CHIRPS incorpora imagens de satélite com dados da estação locais para criar séries temporais de precipitação em grade para análise de tendências e monitoramento de seca sazonal. A resolução temporal de cada pixel é de 1 dia e a resolução espacial é igual a  $0,05^\circ$  de arco, aproximadamente 5,6km [76, 77].

- **Temperatura:** os dados de temperatura foram processados usando o produto da coleção de imagens MOD11A2 V6 que fornece uma temperatura média da superfície da terra de 8 dias por pixel com uma resolução espacial de 1000m em uma grade de 1200x1200km. Cada valor de pixel no MOD11A2 é uma média simples de todos os pixels correspondentes do MOD11A1 LST coletados dentro desse período de 8 dias [78, 79].

Com as séries temporais das variáveis climáticas construímos os GVT e as REAs. A rede de cada variável climática é composta por 417 vértices que se referem aos municípios do estado da Bahia, e as conexões entre eles são definidas de acordo com as suas séries temporais, utilizando o método de sincronização por motivos [5], como indica a Figura 17(d).

Nesse método assumimos motivos de ordem 3, isso implica que 1 motivo corresponde a 45 dias de dados, uma vez que temos coletas quinzenais. Consideramos também,  $Lag(\lambda) = 1$ ,  $Lag\ window=1$  e janelas ( $L_m$ ) compostas por 12 motivos. O Apêndice A apresenta mais detalhadamente a justificativa da escolha desses valores.

Visando uma melhor descrição dos dados das variáveis climáticas, adicionamos um conjunto de novos motivos capazes de descrever os casos em que temos duas medidas iguais e em sequência, ver Figura 19(a).

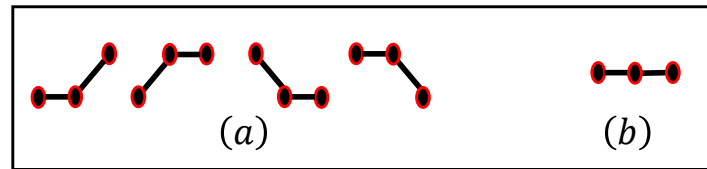


Figura 19 – (a) Padrões de motivos adicionados ao método de sincronização por motivos. (b) Padrão de motivo com 3 medidas seguidas iguais a zero.

Nesse trabalho, no entanto, não consideramos motivos em que três coletas consecutivas são iguais a zero, como o exemplificado na Figura 19(b). Esse padrão poderá ser acrescentado em trabalhos futuros a depender do sistema estudado.

Além dos parâmetros citados, o método de sincronização por motivos define ainda, um parâmetro denominado  $\tau$ , que representa o tempo de atraso entre as séries de motivos. O valor de  $\tau$  para as variáveis climáticas é discutido na seção 2.4.2, quando apresentamos os nossos resultados.

Como indica a Figura 17(h) com os pares de GVT obtidos para as variáveis climáticas, construímos as redes utilizando o método de sincronização de redes multicamadas variantes no tempo.

Na próxima seção apresentaremos os resultados encontrados para esses dados.

## 2.4.2 Resultados

Os resultados dessa aplicação estão divididos em duas partes. Na primeira, subseção 2.4.2.1, definimos os valores dos parâmetros do método de sincronização por motifs [5] e construímos os GVT e a REA para cada uma das variáveis climáticas. E, na segunda parte, subseção 2.4.2.2, apresentamos as redes de sincronização intercadas obtidas para pares de variáveis climáticas utilizando o nosso método de sincronização de redes multicamadas variantes no tempo.

### 2.4.2.1 Redes das Variáveis Individuais

Na subseção anterior apresentamos as variáveis climáticas e as características dos dados utilizados na aplicação. Definimos ainda os valores de alguns dos parâmetros do método de sincronização por motifs, vamos agora definir o valor do parâmetro  $\tau$ .

Como explicado anteriormente, na construção da rede de uma dada variável climática, o parâmetro  $\tau$  caracteriza o tempo de atraso entre as séries de motifs dos municípios. Assim, como 1 motif descreve 3 coletas que são equivalentes a 45 dias, quando consideramos atrasos entre as séries de motifs estamos, por consequência, comparando meses diferentes do ano.

Nessa aplicação obtivemos que, independentemente da variável climática considerada, a maior sincronização entre os municípios ocorre quando não consideramos um tempo de atraso entre os dados, isto é, quando  $\tau = 0$ . Esse valor de  $\tau$  significa que, para uma dada variável, se existir atraso entre os dados dos municípios, ele está em uma escala de tempo menor do que os 45 dias que estamos considerando.

Na Figura 20 ilustramos a situação onde não há atraso e, por consequência, estamos comparando as séries de motifs correspondentes aos mesmos meses para os diferentes municípios.

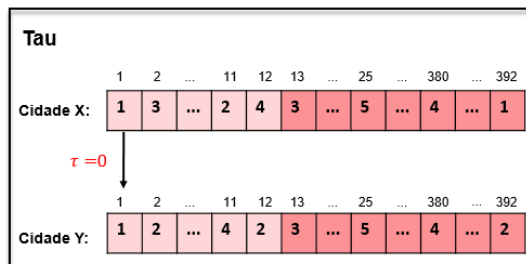


Figura 20 – Vetores com 392 motifs de uma mesma variável para cidades diferentes e dividido em janelas ( $L_m$ ) de 12 motifs cada. Comparações realizadas para tempo de atraso  $\tau = 0$ .

As distribuições da Figura 21 indicam o número de arestas encontradas nos GVT de cada variável climática para diferentes valores de  $\tau$ . Os GVT foram construídos utilizando

os mesmos valores dos parâmetros definidos anteriormente, mas construímos diferentes redes com  $\tau$  variando no intervalo de 0 até 30.

Quanto maior o número de arestas presentes, maior a sincronização entre os componentes da rede. As distribuições mostram que para  $\tau = 0$  obtivemos o maior número de arestas para os GVT em todas as variáveis climáticas, o que significa que a maior sincronização entre as séries das variáveis ocorrem simultaneamente no mesmo período.

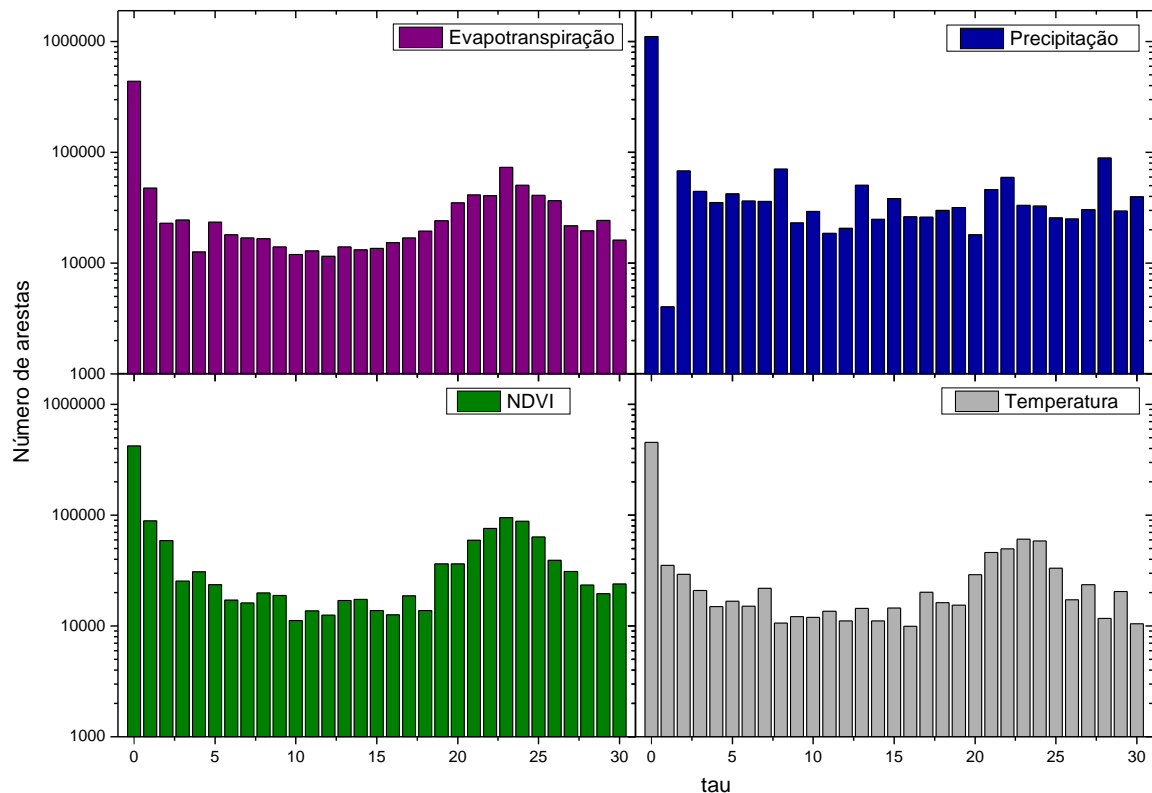


Figura 21 – Número de arestas dos GVT de acordo com o valor do  $\tau$  para cada variável climática.

Para os dados de evapotranspiração, NDVI e temperatura, observamos a presença de ciclos na quantidade de arestas dos GVT em função do valor de  $\tau$ . Nestes gráficos, observamos que, a partir do  $\tau = 1$  a sincronização entre as séries temporais dos municípios diminui, uma vez que o número de arestas presentes nos GVT também diminui. Quando o valor de  $\tau$  está compreendido no intervalo de 20 a 24, completamos um ano de dados, e voltamos a notar um crescimento na sincronização. Isso nos indica que a sincronização dessas variáveis individuais aumenta quando consideramos os dados de um mesmo mês para anos subsequentes.

Para a precipitação, porém, não conseguimos observar a presença de um ciclo único para todos os municípios da Bahia, o que pode ser explicado pelos diferentes regimes de precipitação pluviométrica observados em diferentes regiões do estado [80, 81].

De acordo com a Figura 17(d), definidos os valores dos parâmetros do método de sincronização por motifs [5] construímos os GVT e as REAs, onde os vértices representam os municípios e as arestas a sincronização entre eles para uma determinada variável climática. Para a visualização gráfica das redes e o cálculo dos seus índices utilizamos o software *Gephi - versão 0.9.2* e para os mapas, o software *ArcGIS*. As redes das Figuras 22, 24, 26 e 27 são as REAs das arestas hubs das variáveis climáticas, ou seja, nestas redes consideramos apenas as arestas que possuem um peso maior ou igual a média dos pesos de todas as arestas da rede REA somado dois desvios-padrão,  $\langle w_{i,j} \rangle + 2\sigma_{i,j}$ , como indica a Figura 17(e).

Na Figura 22 temos a REA das arestas hubs para os dados de evapotranspiração durante o intervalo de tempo de 17 anos. Essa rede está sobreposta ao mapa do estado da Bahia com as devidas indicações dos biomas presentes nessa região, como indica a Figura 17(f). A Tabela 4 apresenta os índices de caracterização da rede de evapotranspiração. Como podemos ver, essa rede possui um total de 1821 arestas com 339 vértices conectados. Cada aresta aconteceu em pelo menos 59 ( $\langle w_{i,j} \rangle = 14,35$  e  $\sigma_{i,j} = 20,07$ ) das 392 coletas, ou seja, em 15,05% das coletas, de acordo com o critério discutido anteriormente para uma aresta ser considerada hub.

A dimensão dos vértices na figura é proporcional ao *grau ponderado*. Assim, quanto maior o seu tamanho, maior é a frequência das arestas conectadas a ele durante os GVT. Um vértice com o *grau ponderado* alto indica que este município se conectou com diversos outros municípios ou que ele realizou conexões fortes com os mesmos municípios durante o tempo de vida dos GVT.

Para a evapotranspiração o *grau ponderado* para todo o estado da Bahia varia entre 0 a 6027 (com valor médio igual a 807), esse limite superior provavelmente se encontra na região com maior densidade de arestas localizada nas mesorregiões do vale são-franciscano, extremo oeste e parte do centro-sul bahiano. Neste texto, por uma questão de simplificação, nos referimos a essas regiões apenas como a região centro-oeste. O *grau* varia de 0 a 55 e o *grau médio* da rede é igual a 8,73 o que indica que cada município está conectado, em média, com aproximadamente outros 9 municípios.

Comparamos a rede de evapotranspiração com uma que é composta pelo mesmo número de vértices e arestas, utilizando o modelo de redes aleatórias  $G(N, M)$  de Erdős e Rényi [13]. Todavia, isso foi feito com uma alteração: as arestas recebem pesos aleatórios, sendo eles definidos no mesmo intervalo de peso considerados na rede de evapotranspiração.

O *grau ponderado médio* da rede aleatória é igual 1329, sendo este valor maior que o da rede de evapotranspiração. A justificativa disso, é que a rede aleatória possui apenas uma componente conectada onde os municípios realizam, aproximadamente, o mesmo número de conexões, ocasionando um *grau ponderado* alto e, por consequência, o seu valor médio também. Diferentemente, na rede de evapotranspiração temos uma maior

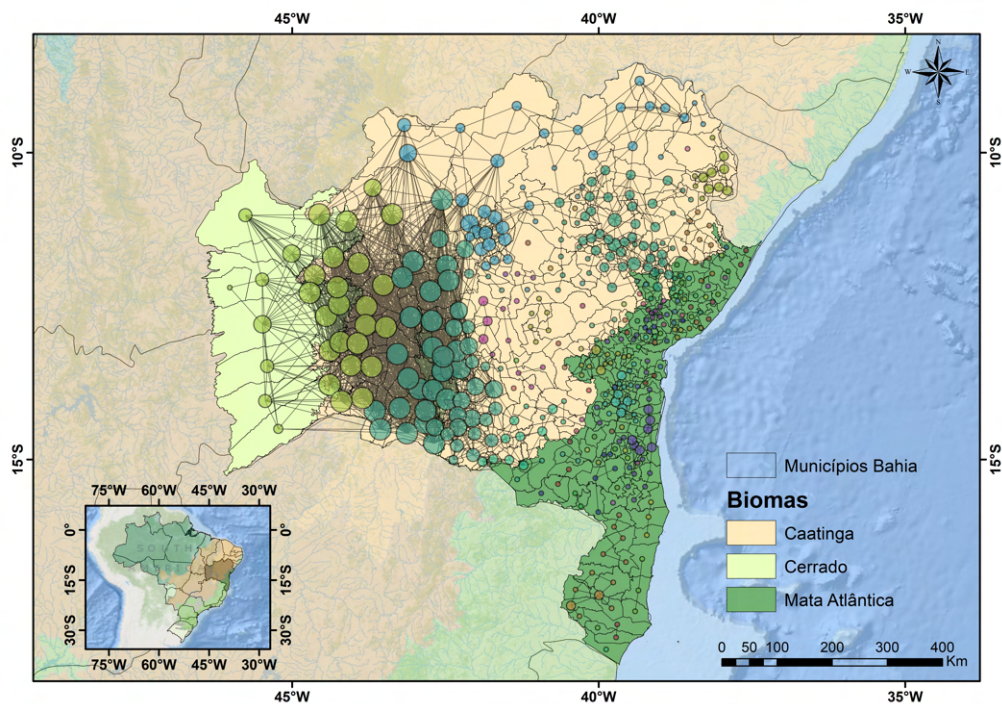


Figura 22 – REA das arestas hubs dos dados de evapotranspiração. Essa rede está sobreposta ao mapa do estado da Bahia, com as devidas especificações dos biomas presentes: Cerrado (lima), Caatinga (amarelo claro), Mata Atlântica (verde).

Índices	Rede dos dados coletados	Rede Aleatória
Vértices	417	417
Arestas	1821	1821
$\langle k \rangle$	8,73(0 a 55)	8,73
$\langle k^\omega \rangle$	807(0 a 6027)	1329
$\langle C \rangle$	0,447	0,026
Modularidade	0,401	0,343

Tabela 4 – Índices das redes dos dados de evapotranspiração e de uma rede similar construída com conexões aleatórias. Os índices  $\langle k \rangle$ ,  $\langle k^\omega \rangle$  e  $\langle C \rangle$  se referem ao grau médio, grau ponderado médio e coeficiente de aglomeração médio, respectivamente. Os intervalos entre parênteses indicam os valores mínimos e máximos para os índices correspondentes.

densidade de arestas apenas em algumas regiões, enquanto em outras os vértices são pouco conectados ou não possuem conexões, gerando assim vértices com *grau ponderado* baixo, se comparado à rede aleatória, ou nulo. Por consequência, o *grau ponderado médio* da rede aleatória é maior do que o da rede de evapotranspiração.

O *coeficiente de aglomeração médio*, que indica o quanto os municípios estão conectados entre si, é igual a 0,447 para a rede de evapotranspiração. Ao considerar a rede aleatória, esse valor é cerca de 17 vezes menor (0,026), indicando que embora os vértices da rede aleatória realizem muitas conexões, eles não formam grupos conectados.

Embora apresentamos índices médios, a rede considerada não é homogênea, mas apresenta uma estrutura dividida em módulos. Neles, os vértices possuem uma maior conexão entre si do que com os demais módulos da rede. Na rede da Figura 22, as cores dos vértices designa os seus diferentes módulos, ao todo temos um total de 114 módulos e *modularidade* igual a 0,401. Para calcular a modularidade das redes apresentadas neste trabalho, utilizamos o método de Louvain [82] implementado no software *Gephi-versão 0.9.2*.

O módulo em que os municípios realizaram o maior número de conexões entre eles, no decorrer do tempo, está representado pela cor verde escuro localizada na região centro-oeste da Bahia. Nela, os vértices possuem *grau ponderado médio* igual a 2160, valor quase 3 vezes maior do que o de toda a rede.

O valor do *grau ponderado médio* da rede de uma determinada variável indica como ocorreu sincronização dos padrões temporais entre os municípios considerados. Quanto maior for o seu valor, maior é a ocorrência da sincronização entre as séries temporais desses municípios.

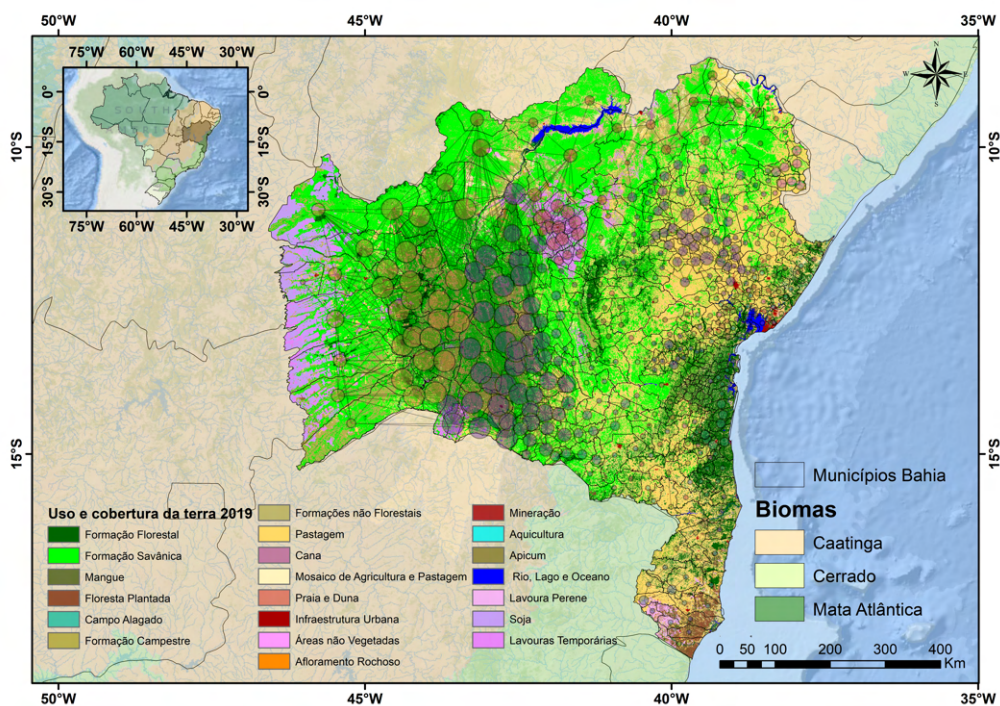


Figura 23 – REA das arestas hubs dos dados de evapotranspiração sobreposta ao mapa da cobertura do solo do estado da Bahia.

Observamos também que muitos municípios não fazem conexões. Neste caso podemos concluir que os padrões temporais desses municípios possuem um nível de semelhança com os demais, menor que o limite inferior de 59 coletas. Há ainda municípios que estão conectados apenas com os vizinhos mais próximos, indicando uma semelhança dos padrões climáticos daquela região. Outros, no entanto, não possuem conexões com os vizinhos



mais próximos o que sugere que nem sempre municípios vizinhos apresentam o mesmo comportamento climático, ainda que façam parte do mesmo bioma.

De acordo com a rede da Figura 22, os municípios da região centro-oeste da Bahia, para os dados de evapotranspiração, apresentam alta densidade de conexões em relação aos municípios das demais regiões. Na Figura 23 temos a rede de evapotranspiração sobreposta ao mapa da cobertura do solo do estado da Bahia, como descreve a Figura 17(g). Como podemos ver, a região centro-oeste é composta majoritariamente pela mesma formação vegetal, savânica. Nesse tipo de formação temos um estrato herbáceo-arbustivo contínuo com a presença de árvores espaçadas. Em outras palavras, a cobertura vegetal dessa região é composta por plantas que crescem pouco, não apresentam caules lenhosos e geralmente possuem um ciclo de vida curto.

Assim como para os dados de evapotranspiração, a região centro-oeste também possui alta densidade de arestas para o NDVI. Esse comportamento pode ser observado na Figura 24, onde temos a REA dos dados de NDVI sobreposta ao mapa do estado da Bahia com as devidas indicações dos biomas presentes nessa região, como indica a Figura 17(f). Na Tabela 5 apresentamos os índices de caracterização encontrados para essa rede.

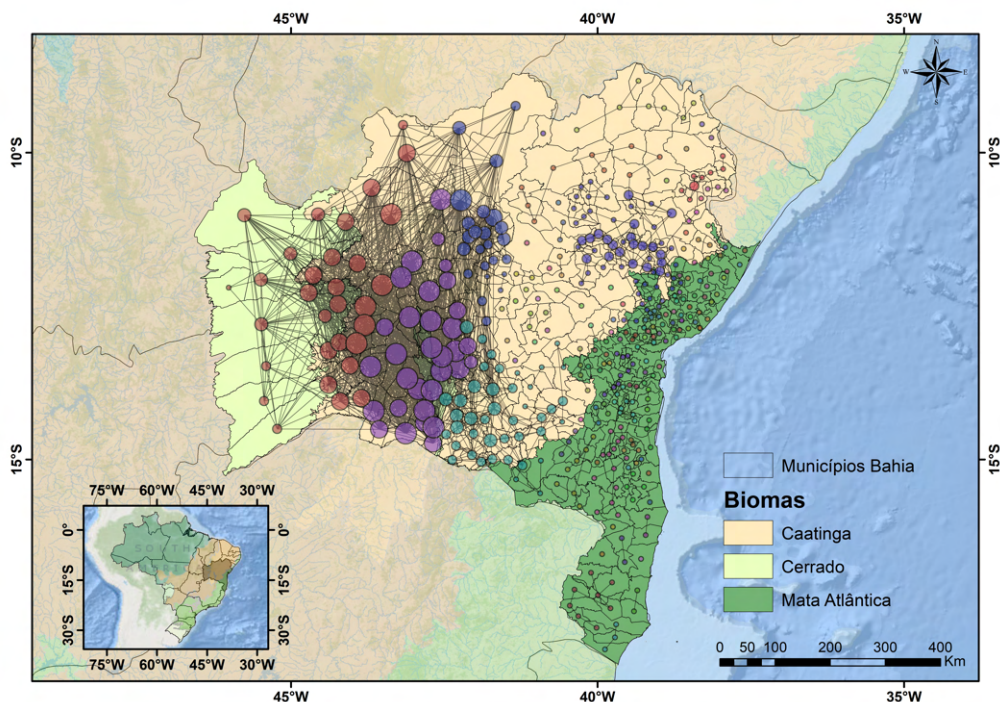


Figura 24 – REA das arestas hubs dos dados de NDVI. Esta rede está sobreposta ao mapa do estado da Bahia, com as devidas especificações dos biomas presentes: Cerrado (lima), Caatinga (amarelo claro), Mata Atlântica (verde).

A rede das arestas hubs de NDVI dispõe de um total de 1780 arestas com 331 vértices conectados. As arestas dessa rede aconteceram em, no mínimo, 59 ( $\langle w_{i,j} \rangle = 15,18$  e  $\sigma_{i,j} = 21,79$ ) das 392 coletas. Quando a frequência de sincronização é abaixo desse limite de 15% das coletas, as arestas não são consideradas na rede e, em alguns municípios

chega-se a ter *grau* igual a 0. Os municípios mais conectados possuem *grau* igual a 59 e *grau médio* igual a 8,5.

O grau ponderado médio é igual a 744 e varia no intervalo de 0 a 5578, indicando a existência de municípios que tiveram alta sincronização entre eles, como por exemplo, o módulo na cor roxa que possui 32 municípios e o *grau ponderado* é igual a 2704, quase 4 vezes mais que o da rede como um todo. O *grau ponderado médio* de uma rede aleatória, construída a partir das mesmas configurações da rede de NDVI, tem valor igual a 1158, e como era esperado, esse valor é maior do que o obtido para a rede dos dados coletados.

Índices	Rede dos Dados	Rede Aleatória
Vértices	417	417
Arestas	1780	1780
$\langle k \rangle$	8,5(0 a 59)	8,5
$\langle k^\omega \rangle$	744(0 a 5578)	1158
$\langle C \rangle$	0,474	0,024
Modularidade	0,416	0,337

Tabela 5 – Índices das redes dos dados de NDVI e de uma rede similar construída com conexões aleatórias. Os índices  $\langle k \rangle$ ,  $\langle k^\omega \rangle$  e  $\langle C \rangle$  se referem ao grau médio, grau ponderado médio e coeficiente de aglomeração médio, respectivamente. Os intervalos entre parênteses indicam os valores mínimos e máximos para os índices correspondentes.

A rede de NDVI possui uma estrutura dividida em 129 módulos, o valor da *modularidade* da rede é igual a 0,416, já para uma rede aleatória esse valor é igual a 0,34. Ainda para a rede de NDVI, o *coeficiente de aglomeração* é igual a 0,474, o que nos indica que os vizinhos de cada município se conectam entre si aproximadamente metade das vezes que poderiam se conectar. Em uma rede aleatória, a probabilidade dos municípios conectados a um determinado município estarem conectados entre si é aproximadamente 20 vezes menor, 0,024.

A Figura 25 exibe a REA das arestas hubs dos dados de NDVI sobreposta ao mapa de uso e cobertura da terra, como sugere a Figura 17(g). Nessa rede temos que os módulos mais conectados estão na região centro-oeste, onde a cobertura do solo é predominantemente a formação savânica.

Em síntese, temos que, as redes de evapotranspiração e NDVI obtidas pelo método de sincronização por motifs são definidas tanto pela presença de muitos municípios com poucas ou sem conexões na região da Mata Atlântica e uma parcela da Caatinga, quanto por módulos superconectados na região centro-oeste da Bahia.

A variável evapotranspiração, como vimos, está fortemente ligada com a evaporação da água das superfícies de rios e de solos úmidos e também com a transpiração das plantas, e a variável NDVI está relacionada com a atividade fotossintética das plantas. Dessa forma,

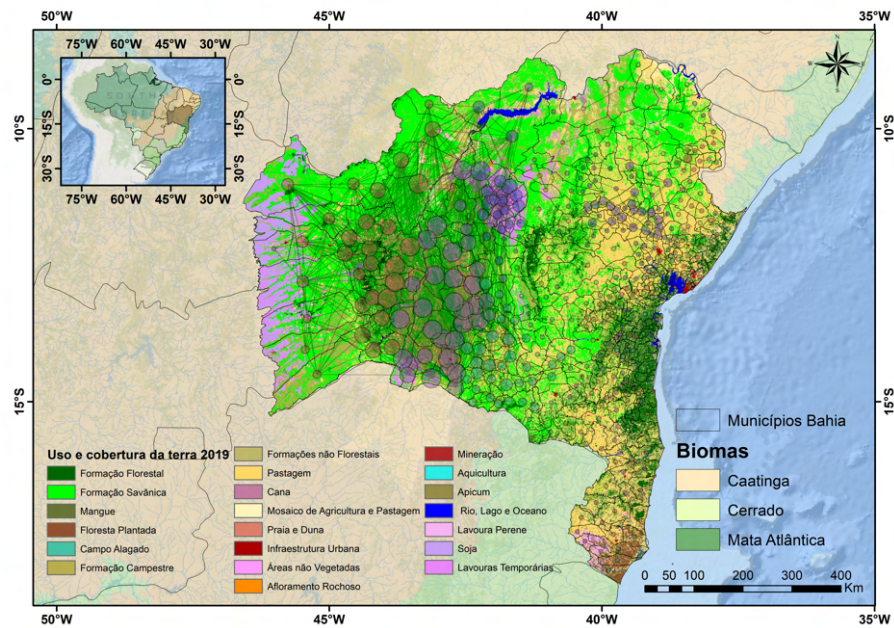


Figura 25 – REA das arestas hubs dos dados de NDVI sobreposta ao mapa da cobertura do solo do estado da Bahia.

supomos que os módulos superconectados destas redes se devem à contribuição de dois fatores: primeiro, a localização da Bacia hidrográfica do Rio São Francisco que corta essa região e, segundo, a sua forte atividade agrícola que faz com que ela seja a maior área de expansão agrícola irrigada do estado da Bahia com o cultivo de produtos, como: algodão, feijão, milho, soja e alguns cultivos frutíferos. Nessas áreas, a disponibilidade de água durante todo o ano favorece uma maior atividade fotossintética das culturas agrícolas.

Na Figura 26 exibimos a REA das arestas hubs para os dados de precipitação sobreposta ao mapa dos biomas do estado da Bahia, como indica a Figura 17(f). A Tabela 6 apresenta os índices de caracterização dessa rede. No seu processo de construção todas as arestas que apareceram em número menor que 62 coletas ( $\langle w_{i,j} \rangle = 18,24$  e  $\sigma_{i,j} = 21,75$ ) foram descartadas, restando assim, um total de 3110 arestas com 413 vértices conectados. Os municípios dessa rede possuem *grau* variando de 0 a 55, com *grau médio* igual a 14,9.

Já o *grau ponderado*, que indica a intensidade das conexões que um determinado município realiza, varia no intervalo de 0 a 5032 com valor médio igual a 1390. Uma rede aleatória com as mesmas configurações apresenta o *grau ponderado* no intervalo de 844 a 4608 e média igual a 2393. Esta relação já foi obtida nas duas redes anteriores e entendemos que o valor do *grau médio* é maior, pois as arestas estão distribuídas uniformemente na rede.

A rede de precipitação é dividida em 14 módulos com *modularidade* igual a 0,678, o *coeficiente de aglomeração médio* é igual a 0,543 indicando a existência de agrupamentos de municípios conectados entre si. Esse é o maior valor do *coeficiente de aglomeração médio* encontrado para as REAs das arestas hubs das 4 variáveis climáticas consideradas. Para

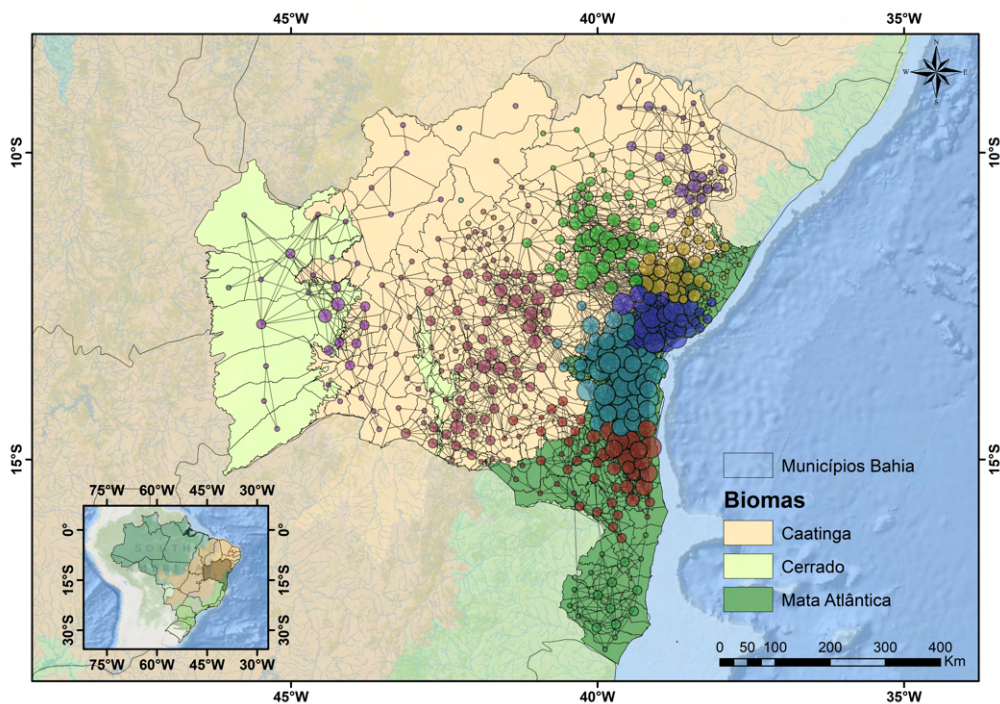


Figura 26 – REA das arestas hubs dos dados de precipitação. Esta rede está sobreposta ao mapa do estado da Bahia, com as devidas especificações dos biomas presentes: Cerrado (lima), Caatinga (amarelo claro), Mata Atlântica (verde).

Índices	Rede dos Dados	Rede Aleatória
Vértices	417	417
Arestas	3110	3110
$\langle k \rangle$	14,9(0 a 55)	14,9
$\langle k^{\omega} \rangle$	1390(0 a 5032)	2393
$\langle C \rangle$	0,543	0,04
Modularidade	0,678	0,236

Tabela 6 – Índices das redes dos dados de precipitação e de uma rede similar construída com conexões aleatórias. Os índices  $\langle k \rangle$ ,  $\langle k^{\omega} \rangle$  e  $\langle C \rangle$  se referem ao grau médio, grau ponderado médio e coeficiente de aglomeração médio, respectivamente. Os intervalos entre parênteses indicam os valores mínimos e máximos para os índices correspondentes.

uma rede aleatória com as mesmas configurações da rede das arestas hubs de precipitação, a possibilidade de encontrar municípios em que os seus vizinhos estejam conectados entre si é cerca de 13 vezes menor, uma vez que o *coeficiente de aglomeração* dessa rede é igual a 0,04.

O módulo com maior frequência de sincronização entre os municípios e, por consequência, com maior *grau ponderado* está localizado na Mata Atlântica, módulo na cor verde-azulado (■) composto por 56 municípios conectados por 681 arestas. Esse módulo tem *grau ponderado* igual a 2334 e representa quase o dobro do *grau ponderado* para toda a REA hub de precipitação.

Diferentemente dos padrões anteriores, observamos agora que as sincronizações ocorreram em maior parte na região da Mata Atlântica e, por consequência, no litoral e suas intermediações. Este resultado já era esperado, visto que esta região possui clima tropical quente e úmido com índices pluviométricos anuais superiores a 1500mm, diferentemente das demais regiões marcadas por períodos de seca.

Para o interior da Bahia, exceto para a parte norte e o vale do rio São Francisco, também observamos a presença de módulos, porém com sincronização menor do que as encontradas para a região da Mata Atlântica. Acreditamos que este comportamento também está relacionada ao clima dessa região, tropical úmido e seco. Este tipo de clima possui duas estações: chuvosa no verão e seca no inverno. E, como vimos, os ciclos de precipitação para cada município não são necessariamente os mesmos, o que sugere uma menor sincronização entre estes municípios.

Nas regiões do norte do estado e no vale do São Francisco, o clima é semiárido com pluviosidade média anual inferior a 700mm com longos períodos de estiagem. Os municípios dessa região não formam módulos superconectados e alguns não realizam conexões. Isto porque, na construção dessas redes não são considerados os períodos de estiagem, ou seja, não há um motif associado a três valores em sequência iguais a zero, como foi mencionado na seção 2.4.2.1.

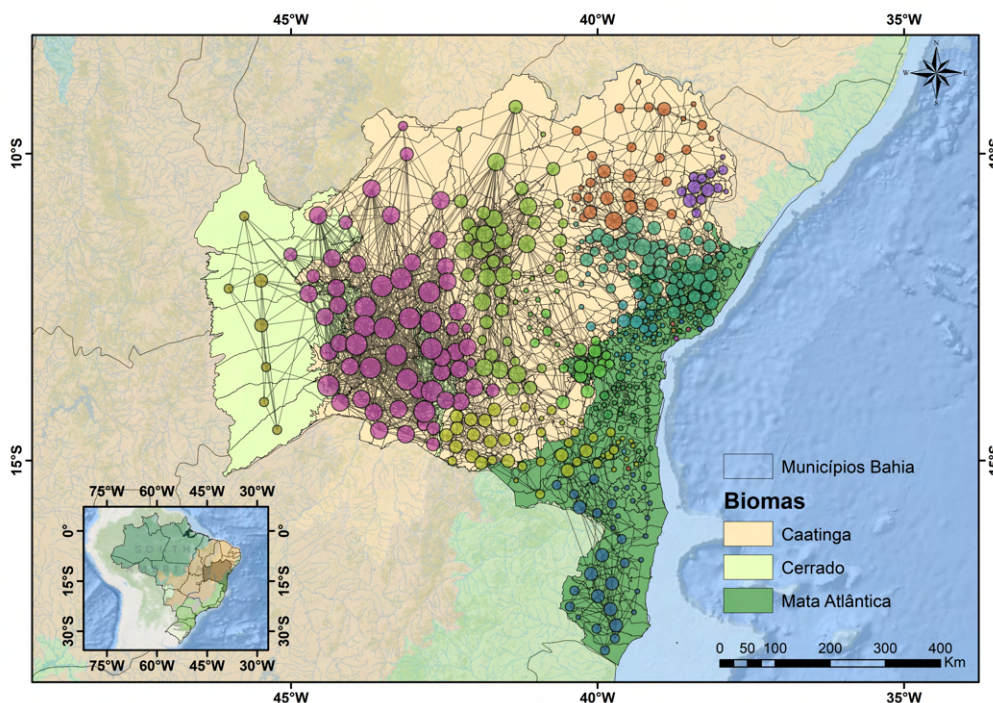


Figura 27 – REA das arestas hubs dos dados de temperatura. Essa rede está sobreposta ao mapa do estado da Bahia, com as devidas especificações dos biomas presentes: Cerrado (lima), Caatinga (amarelo claro), Mata Atlântica (verde).

E, por fim, a última REA das arestas hubs a ser analisada é a referente aos dados de temperatura. A Figura 27 apresenta essa rede sobreposta ao mapa do estado da Bahia

com os biomas dessa região, como indica a Figura 17(f). A Tabela 7 apresenta os índices de caracterização da rede de temperatura. Essa rede é composta por 2311 arestas com 408 vértices conectados. As arestas dessa rede apareceram em pelo menos 35 coletas,  $\langle w_{i,j} \rangle = 10, 16$  e  $\sigma_{i,j} = 12, 30$ . O *grau médio* da rede é igual a 11 e varia no intervalo de 0 a 46, ou seja, os municípios mais conectados fazem, no máximo, 46 ligações.

O *grau ponderado* está compreendido entre o intervalo de 0 a 2292, com *grau ponderado médio* igual a 600. Comparado a uma rede aleatória, esse índice é igual a 1067.

Essa rede é fragmentada em 20 módulos com *modularidade* igual a 0,682. O módulo em que houve uma maior sincronização entre as arestas e, por consequência, o maior valor do *grau ponderado médio* foi o módulo em roxo, composto por 61 municípios e 754 arestas. O *grau ponderado médio* assumiu o valor de 1428, aproximadamente 2,4 vezes maior do que o valor do *grau ponderado médio* da rede de temperatura para toda a Bahia.

O *coeficiente de aglomeração médio* entre os municípios dessa rede é igual a 0,458, o que demonstra uma tendência dos vizinhos de um município serem vizinhos entre si. Em uma rede aleatória, o valor do *coeficiente de aglomeração* é igual a 0,029, isto é, cerca de 15,8 vezes menor, o que reforça que os agrupamentos encontrados na rede de temperatura não são frutos do acaso.

Índices	Rede dos Dados	Rede Aleatória
Vértices	417	417
Arestas	2311	2311
$\langle k \rangle$	11(0 a 46)	11
$\langle k^\omega \rangle$	600(0 a 2292)	1067
$\langle C \rangle$	0,458	0,029
Modularidade	0,682	0,292

Tabela 7 – Índices das redes dos dados de temperatura e de uma rede similar construída com conexões aleatórias. Os índices  $\langle k \rangle$ ,  $\langle k^\omega \rangle$  e  $\langle C \rangle$  se referem ao grau médio, grau ponderado médio e coeficiente de aglomeração médio, respectivamente. Os intervalos entre parênteses indicam os valores mínimos e máximos para os índices correspondentes.

Para essas análises é importante considerar que o clima do estado da Bahia é o clima tropical. Este tipo de clima é caracterizado por temperaturas elevadas e baixa amplitude térmica, o que significa que a temperatura da região não tem grandes variações nem durante o dia, nem ao longo do ano. Em vista disso, a rede encontrada para estes dados apresenta uma distribuição espacial uniforme, onde praticamente todos os municípios estão conectados entre si com diferentes intensidades de sincronização.

Outro fator que deve ser ressaltado é que a distribuição de temperatura de uma região está associada a sua topografia, sendo que para o estado da Bahia as temperaturas mais altas se encontram no centro-oeste acompanhando o vale do Rio São Francisco. Em

paralelo, a rede obtida para a temperatura apresenta módulos superconectados localizados também na região centro-oeste. É provável que a semelhança na distribuição de temperatura dos municípios dessa região tenha gerado os módulos superconectados encontrados na rede de temperatura.

#### 2.4.2.2 Redes de Sincronização para Pares de Variáveis Climáticas

A partir daqui iremos analisar as redes de Incidência-Fidelidade derivadas do nosso modelo de sincronização de redes multicamadas variantes no tempo, como prevê a Figura 17(h). No estudo da sincronização intercamadas entre as diferentes variáveis climáticas dos municípios do estado da Bahia consideramos os seguintes pares de GVT: evapotranspiração - NDVI, evapotranspiração - precipitação, evapotranspiração - temperatura e NDVI - precipitação.

As redes de Incidência-Fidelidade obtidas passaram por um processo de filtragem diferente do adotado nas REAs da subseção anterior. Nas redes de Incidência-Fidelidade utilizamos GVT multicamadas construídos a partir do modelo de redes aleatórias de Erdős e Rényi  $G(N, M)$ , para filtrar as arestas significativas das redes de Incidência-Fidelidade dos dados coletados. Enquanto que nas REAs, a definição das arestas significativas considera as médias e os desvios padrões dos pesos das arestas.

Para construir os GVT aleatórios consideramos as mesmas configurações dos GVT dos dados coletados: mesmo tempo de vida, mesmo número de vértices e mesmo número de interconexões em cada instante de tempo. A partir desses GVT aleatórios, calculamos a Incidência-Fidelidade correspondente aos mesmos pares das variáveis climáticas consideradas. O valor máximo encontrado define o limiar que nos indica se as arestas da rede de Incidência-Fidelidade dos dados coletados são significativas.

Assim, para um determinado par de variáveis climáticas, quando o valor da Incidência-Fidelidade de uma aresta for maior que o limiar correspondente, a aresta permanece na rede de Incidência-Fidelidade dos dados observados. Caso contrário, ela passa a ser tida como fruto do acaso e é removida da rede. Esse processo é indicado no fluxograma apresentado na subseção 2.4.1, especificamente Figura 17(i).

Esse processo é semelhante ao adotado no método de sincronização por motivos [5]. No entanto, as redes obtidas pelo nosso método considera um limiar de 100%, enquanto que no método de sincronização por motivos o limiar considerado é de 95%. A justificativa para essa diferença é que as redes do nosso método são muito conectadas e não conseguimos observar o padrão de conexão quando consideramos o mesmo limiar utilizado no método de sincronização por motivos.

As redes de Incidência-Fidelidade das Figuras 28 e 29 já passaram pelo processo de filtragem e são referentes as combinações das variáveis evapotranspiração - NDVI e evapo-

transpiração - temperatura, respectivamente. Essas redes são prenunciadas na Figura 17(j). Uma aresta na rede de Incidência-Fidelidade de um par de variáveis climáticas quaisquer, indica que quando aconteceu sincronização entre dois municípios para uma dada variável climática, é possível que tenha acontecido também para a outra, indicando uma afinidade climática para esses municípios.

A análise das redes de Incidência-Fidelidade dos pares evapotranspiração - NDVI e evapotranspiração - temperatura será feita em conjunto, uma vez que elas obedecem o mesmo padrão de conexão. Para essas combinações, podemos notar uma maior densidade de arestas na região centro-oeste, o que sugere uma maior sincronização das variáveis climáticas nessa região do que entre as demais regiões.

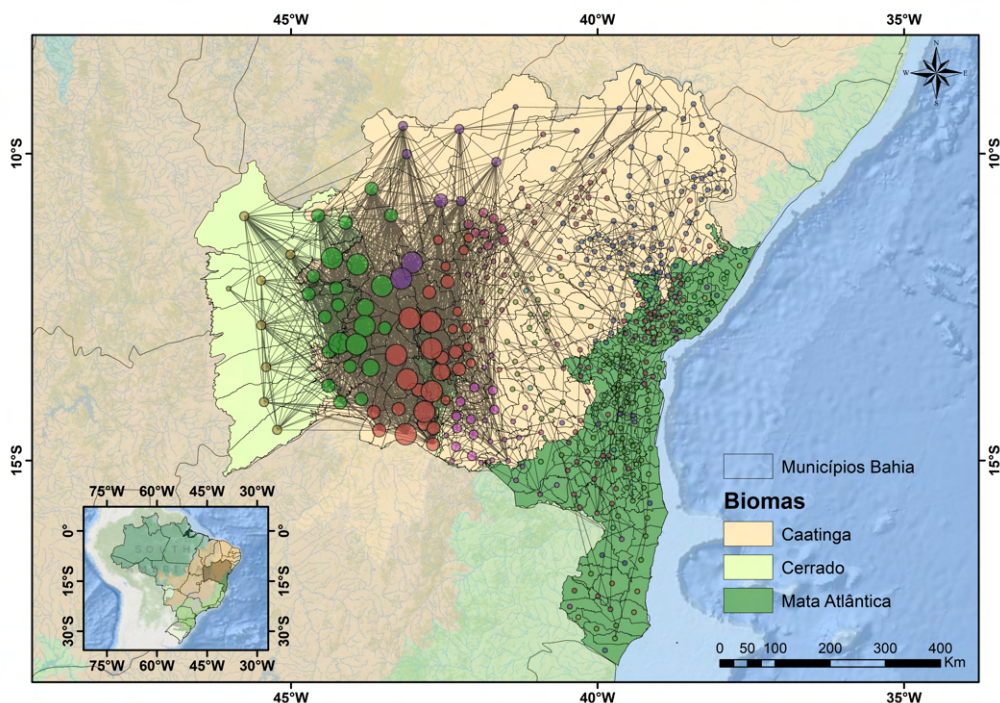


Figura 28 – Rede de Incidência-Fidelidade do par evapotranspiração - NDVI.

Quando analisamos individualmente as REAs das arestas hubs da evapotranspiração e do NDVI, confirmamos que a região centro-oeste apresenta uma quantidade maior de conexões do que as demais regiões, como mostrado nas Figuras 22 e 24. De forma análoga, para os dados de temperatura, embora a REA das arestas hubs seja espacialmente distribuída, como mostra a Figura 27, ela também apresenta módulos com maior *grau ponderado* na região centro-oeste.

Assim como as REAs das arestas hubs das variáveis evapotranspiração, NDVI e temperatura, as redes de Incidência-Fidelidade resultantes da combinação dessas variáveis também apresentam alta densidade de arestas na região centro-oeste em relação as demais regiões da Bahia. Isto nos indica que as conexões presentes nas redes individuais acontecem de forma simultânea para as diferentes variáveis, ao considerar não somente a presença



simultânea em relação ao tempo, mas também em relação a frequência que as conexões aconteceram individualmente.

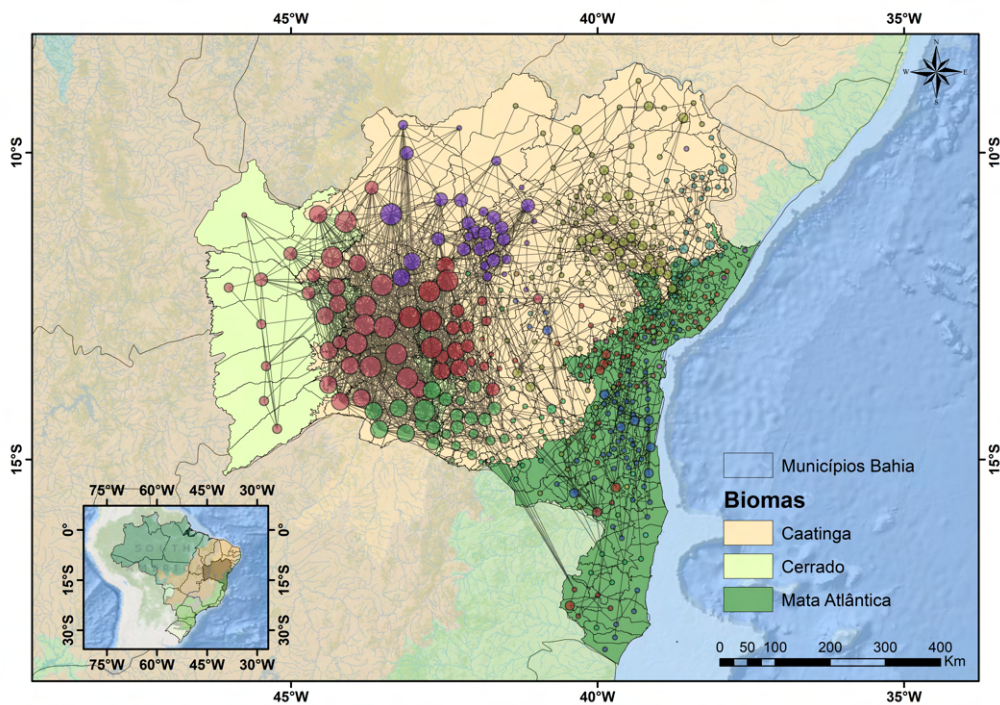


Figura 29 – Rede de Incidência-Fidelidade do par evapotranspiração - temperatura.

Todas as redes de Incidência-Fidelidade obtidas pelo nosso método para essa aplicação exibem ligações, predominantemente, de curta distância. Isso significa que há uma propensão das regiões vizinhas apresentarem o mesmo comportamento de conexão entre diferentes variáveis.

No entanto, observamos também a presença de ligações de longa distância, tanto para municípios do mesmo bioma, quanto para diferentes biomas. Essas ligações não existiam nas redes das arestas hubs individuais analisadas anteriormente, inclusive observamos conexões realizadas por municípios que antes estavam desconectados. Estas ligações são atribuídas ao índice Fidelidade e indicam que arestas de regiões e variáveis diferentes podem estar sincronizadas ao se considerar apenas os instantes em que pelo menos uma delas apareceu em alguma das camadas.

É importante reafirmar que no cálculo da Fidelidade todas as arestas são consideradas, incluindo as que foram eliminadas no processo de filtragem da REA das arestas hubs. Isto porque o modelo de sincronização de redes multicamadas considera os GVT e não a REA hub. As ligações de longa distância são frutos dessas arestas que apareceram abaixo do limite de filtragem, mas que exibem forte sincronização no que se refere a frequência em que elas acontecem.

Entendemos que a rede de Incidência-Fidelidade das variáveis evapotranspiração - NDVI reflete a relação de dependência que existe na definição dessas variáveis. Na região

centro-oeste, por exemplo, as medidas quinzenais de evapotranspiração e NDVI evoluem obedecendo o mesmo padrão de comportamento, ou seja, se duas cidades se conectam na rede da variável evapotranspiração, provavelmente também se conectam na rede da variável NDVI, no mesmo intervalo de tempo.

Para essas variáveis, as sincronizações de longa distância evidenciadas pelo índice Fidelidade acontecem, em maior frequência, nos períodos de secas simultâneas de dois biomas diferentes. Nesses casos, a atividade fotossintética das plantas diminuem em ambas regiões, e assim os valores do NDVI caem. De forma análoga, com a estiagem e a redução da atividade fotossintética das plantas, a evapotranspiração também sofre uma redução, obedecendo o mesmo padrão de decrescimento que a evapotranspiração e, ao mesmo tempo, sincronizando essas variáveis.

Como já mencionado, a rede de Incidência-Fidelidade do par evapotranspiração-temperatura também apresenta módulos super conectados localizados na região centro-oeste, que é a região com maior sincronização para as redes individuais de evapotranspiração e de temperatura. A região centro-oeste faz parte do semi-árido e, por isso, apresenta baixos índices pluviométricos e altas temperaturas durante todo o ano. Acreditamos que a maior contribuição para a sincronização nessa região, se deve, principalmente, ao índice Incidência.

A seguir iremos avaliar a rede de Incidência-Fidelidade em que as redes das arestas hubs das variáveis climáticas individuais não possuem módulos superconectados na mesma região. Assim, vamos considerar as redes de Incidência-Fidelidade das Figuras 30 e 31 que são referentes aos pares de variáveis: evapotranspiração - precipitação e NDVI - precipitação. Essas redes estão sobrepostas ao mapa dos biomas do estado da Bahia e são previstas no último passo do fluxograma, Figura 17(j).

Para as redes individuais que compõem as combinações consideradas neste momento, vimos que a rede das arestas hubs da variável evapotranspiração apresenta uma maior sincronização na região centro-oeste, envolvendo o Cerrado e parte da Caatinga e que esse mesmo comportamento é observado para as redes hubs da variável climática NDVI. Ainda na subseção 2.4.2.1 discutimos sobre a rede de precipitação. Vimos que ela apresenta uma maior sincronização para os municípios da região da Mata Atlântica. Para as demais regiões observamos municípios sem ou com conexões apenas com os vizinhos mais próximos, inclusive na região centro-oeste.

As redes de Incidência-Fidelidade resultantes dos pares das variáveis climáticas evapotranspiração-precipitação e NDVI-precipitação, apresentam padrões de conexão espacialmente mais distribuídos do que os encontrados nas REAs hubs individuais. Essa configuração está atrelada ao fato que os módulos mais conectados de cada REA individual pertencem a diferentes regiões do estado. Assim, as conexões que aparecem com frequência alta para uma dada variável climática numa determinada região, não aparecem neces-

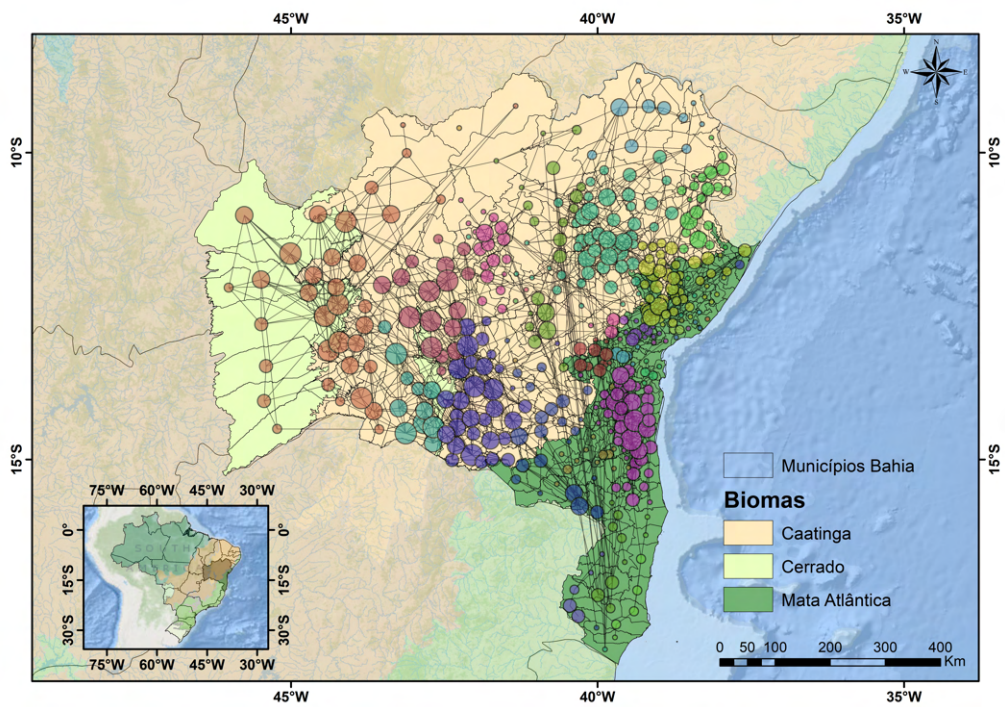


Figura 30 – Rede de Incidência-Fidelidade do par evapotranspiração - precipitação.

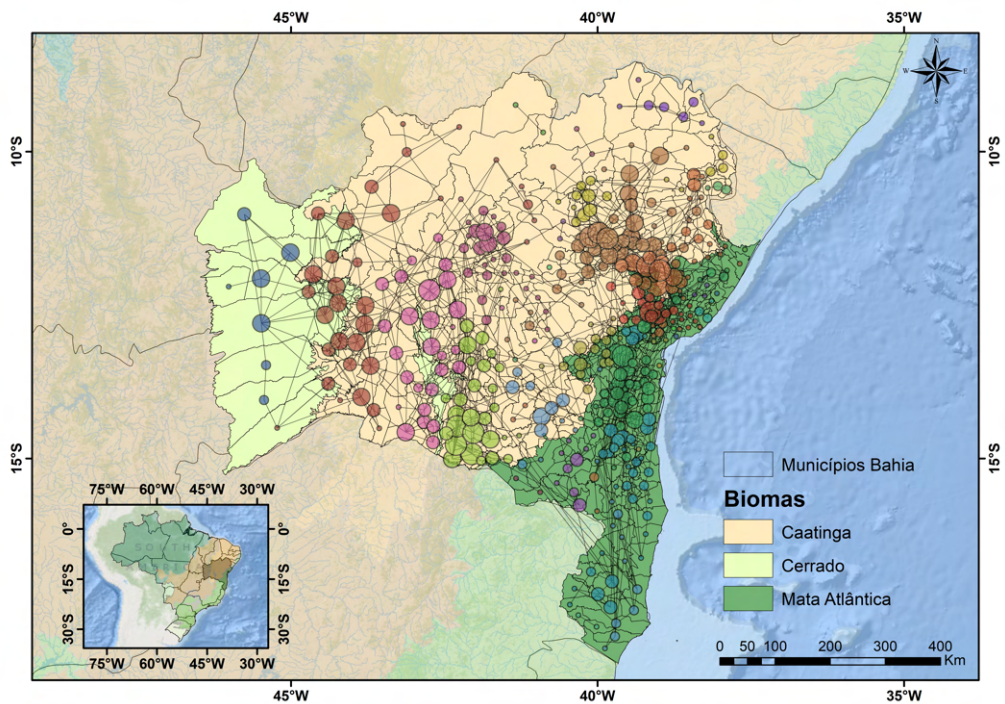


Figura 31 – Rede de Incidência-Fidelidade o par NDVI - precipitação.

sariamente de forma simultânea para a outra variável, nessa mesma região. Esperamos que exista uma dependência do número de arestas das redes individuais com o resultado da rede de Incidência-Fidelidade. Pretendemos avaliar esse comportamento em trabalhos futuros.

Para a rede de Incidência-Fidelidade gerada da sincronização do par de variáveis evapotranspiração-precipitação (ver Figura 30), também observamos a presença de conexões de curta e longa distância. Embora a definição de evapotranspiração tenha relação direta com a precipitação, temos que considerar também o NDVI. Em vista disso, a rede de Incidência-Fidelidade dessas variáveis apresenta padrões de sincronização distintos das redes individuais de cada variável.

Como já discutido, a REA das arestas hubs de evapotranspiração apresenta módulos com alta sincronização na região centro-oeste. Por outro lado, quando analisamos as séries temporais da precipitação, observamos que os poucos módulos encontrados nessa região estão relacionados aos baixos índices de precipitação e aos períodos de estiagem. Assim, a rede de Incidência-Fidelidade destas variáveis, para a região centro-oeste, não apresenta módulos superconectados. O mesmo acontece na região da Mata Atlântica, onde temos módulos superconectados para a variável precipitação, devido a frequente presença de chuvas, mas não para a evapotranspiração. O resultado deste processo de sincronização é o aparecimento de pequenos módulos com ligações de curta distância.

Temos ainda que as ligações de longa distância encontradas nessa rede se devem aos poucos períodos simultâneos de chuva intensa entre regiões diferentes. Nesses períodos, os índices de precipitação do Cerrado e da Caatinga, por exemplo, se aproximam dos valores obtidos para a Mata Atlântica.

Por fim, a rede de Incidência-Fidelidade das variáveis NDVI-precipitação da Figura 31, apresenta uma rede com módulos distribuídos na região centro-oeste, nordeste e na Mata Atlântica. Como vimos, a região centro-oeste é reconhecida como uma região voltada para a agricultura variando entre tipos de cultivos durante todo ano. Como os períodos de chuva desta região estão acumulados em determinados meses do ano, acreditamos que a sincronização ocorra apenas nesses períodos. Essa explicação se estende ao módulo encontrado na região nordeste. Para os módulos da Mata Atlântica temos chuva com frequência diária, os módulos encontrados nessa região são resultados da fragmentação da Mata Atlântica e os altos índices de chuva.

Nessa aplicação analisamos as redes de Incidência-Fidelidade de um par de variáveis climáticas, considerando as seguintes possibilidades: quando as REAs das arestas hubs de cada variável possuem módulos super conectados localizados na mesma região ou em regiões diferentes. Vimos que as redes obtidas apresentam predominantemente ligações de curta distância, mas notamos também a presença de ligações de longa distância que não eram observadas nas redes das arestas hubs individuais. Por fim, entendemos, ainda, a contribuição dos índices de Incidência e Fidelidade para a construção das redes resultantes do método de sincronização de redes multicamadas variantes no tempo proposto nesse trabalho.

## 2.5 Conclusões

Neste trabalho propomos um método de sincronização intercamadas que visa calcular a sincronização entre dois sistemas que possuem algum nível de conexão entre eles e que as suas conexões internas mudam no decorrer do tempo. Para isso, consideramos alguns conceitos já estabelecidos na literatura, como as redes multicamadas, as redes variantes no tempo e o índice Incidência-Fidelidade para redes de palavras.

A definição original desse índice considera redes com apenas uma camada estática. Enquanto que no nosso método, redefinimos o índice Incidência-Fidelidade para explorar a sincronização entre redes com duas camadas, cujas conexões mudam com o tempo, o que nos possibilita uma série de aplicação. A validação desse método foi realizada tanto para modelos nulos, quanto para dados com acoplamento preestabelecido.

Na sua aplicação consideramos dados de evapotranspiração, NDVI, precipitação e temperatura de todo o estado da Bahia, em um período de 17 anos. Na construção das redes individuais dessas variáveis, verificamos a presença de ciclos nas distribuições do número de arestas para diferentes tempos de atraso ( $\tau$ ). Estes ciclos, no entanto, não foram observados para os dados de precipitação ao considerar todos os municípios do estado. Isto nos sugere que não existe um ciclo único para todos os municípios, o que já foi observado anteriormente em outros trabalhos com uma abordagem climatológica.

As redes de sincronização intercamadas obtidas para essas variáveis climáticas trazem as seguintes conclusões gerais: em todas as redes de Incidência-Fidelidade observamos a existência de ligações de longa distância, relacionadas ao índice Fidelidade, mas as conexões predominantes foram as conexões de curta distância. Estas, sugerem que há uma predisposição de municípios vizinhos apresentarem padrões de conexão semelhantes entre as diferentes variáveis climáticas consideradas.

Observamos ainda que, quando consideramos pares de variáveis em que as REAs das arestas hubs possuem módulos superconectados localizados em regiões distintas, a rede de Incidência-Fidelidade obtida é formada por conexões espacialmente distribuídas. Nesse caso, o índice Incidência oferece uma contribuição pequena na sincronização das redes, visto que, os módulos superconectados ocorrem em regiões diferentes, diminuindo assim a probabilidade das arestas coocorrerem. Do contrário, quando consideramos pares de variáveis em que as REAs das arestas hubs possuem módulos superconectados localizados nas mesmas regiões, a rede de Incidência-Fidelidade resultante também apresenta esta mesma configuração. Neste caso, a Incidência exerce grande influência na rede de sincronização.

O método de sincronização multicamadas aplicado aos dados de variáveis climáticas, nos permite análises, como: avaliar do ponto de vista climático inúmeras variáveis ao mesmo tempo e observar exatamente como elas se comportam, como funcionam e variam no tempo. Podemos testar cenários, estudando o impacto das variáveis na sincronização,

ou ainda, adicionar tempos de atrasos entre as variáveis.

Em síntese, o método que desenvolvemos é adequado para investigar sistemas que podem ser analisados sob diferentes parâmetros e que as conexões referentes a esses parâmetros mudam no decorrer do tempo. O resultado desse método é uma rede com conexões de curto alcance, influenciadas principalmente pela Incidência e de longo alcance que são relacionadas com a Fidelidade.

O método de sincronização de redes multicamadas variantes no tempo tem ampla capacidade de ser aplicado em diversos sistemas, dentre eles, podemos citar sistemas voltados para a área da neurociência ou da epidemiologia.

# Apêndices

# APÊNDICE A – Construção das Redes utilizando o Método de Sincronização por Motifs

Como vimos anteriormente, cada uma das 417 cidades da Bahia possui um total de 4 séries temporais que são referentes as variáveis climáticas consideradas. Estas séries têm em média 392 pontos, que são as medidas que ocorrem quinzenalmente durante 17 anos. De acordo com o método de sincronização por motifs [5] as séries temporais são reescritas como uma sequência de motifs, devido a natureza dos dados, consideramos que o *Lag* ( $\lambda$ ) é igual a 1, ou seja, pegamos pontos seguidos da série como exemplifica a Figura 32.

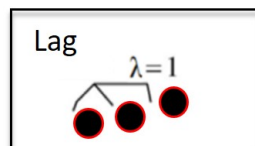


Figura 32 – Exemplo de um motif construído com  $\lambda = 1$

Consideramos que os motifs são de ordem 3, isto é, formados por três pontos da série. Assim, para os dados das variáveis climáticas, um motif representa uma sequência de 3 medidas coletadas num período médio de 45 dias. A Figura 33 ilustra uma série temporal de um período de 1 ano e 5 meses, cada círculo representa um ponto da série, as letras representam a inicial do mês que aquele ponto pertence. Uma sequência de 3 pontos designa um motif, assim o primeiro motif é a sequência *JJF*, o segundo *JFF* e assim por diante.

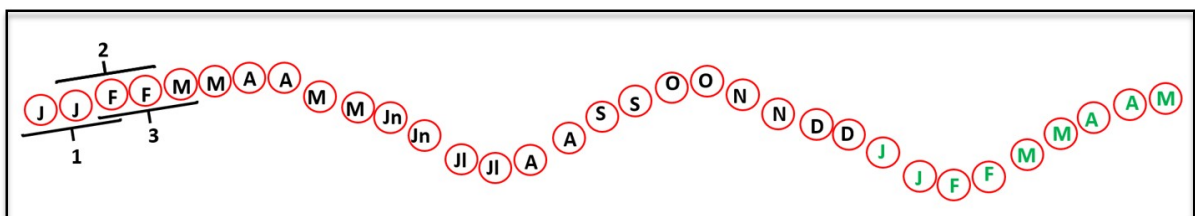


Figura 33 – Exemplo ilustrativo de uma série temporal de um período de 1 ano e 5 meses.

Após a decodificação das séries temporais utilizando os motifs padrões definidos por Rosário e colaboradores [5], construímos os vetores de motifs que descrevem essas séries como exemplifica a Figura 34(a). Para estes dados definimos o tamanho da janela ( $L_m$ ) igual à 12 motifs que corresponde a uma sequência de 8 meses.



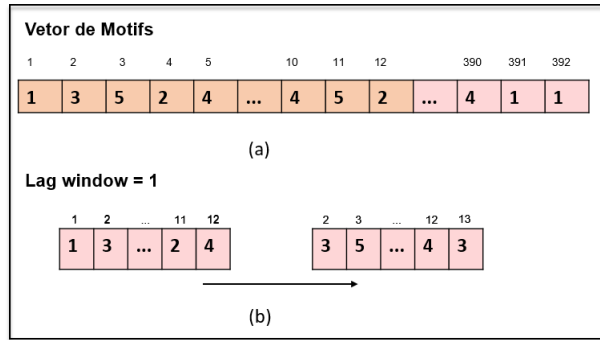


Figura 34 – (a) Série temporal reescrita como um vetor de motifs. (b) Exemplo do movimento das janelas para *Lag window* = 1.

E por fim, na construção das redes consideramos o *Lag window* igual a 1, como mostra a 34(b), o que significa que para passar de uma janela para a seguinte, corremos apenas uma posição no vetor de motifs.

# APÊNDICE B – Biomas presentes no estado da Bahia

A seguir, apresentamos algumas características gerais sobre os biomas da Bahia. Essas informações foram obtidas na referência [83].

## **Caatinga**

O bioma Caatinga ocupa 54% do território baiano, trata-se de um bioma puramente brasileiro e predominante no Nordeste do Brasil. O seu nome é de origem tupi-guarani e significa “mata ou floresta branca”, isto porque a maioria das plantas desse tipo de ecossistema, perde as folhas no período da seca, prevalecendo na paisagem um aspecto claro e esbranquiçado dos troncos das árvores. Sem as folhas verdes, a taxa fotossintética é drasticamente reduzida e as plantas entram em estágio de economia de energia e uso das reservas de água. Nos meses de chuva a paisagem assume variados tons de verde com a rebrota das folhas das árvores e com o surgimento de diversas plantas logo nas primeiras chuvas.

O clima da Caatinga é o Semiárido, o que significa que essa região apresenta altas temperaturas, chuvas escassas e mal distribuídas com longos períodos de estiagem, os níveis pluviométricos atingem, em média, cerca de 800 mm/ano. Os períodos chuvosos, possuem duração de 3 a 5 meses, geralmente de janeiro a maio, e os níveis pluviométricos podem chegar até os 1000 mm/ano. O período de seca ocorre, na maior parte do ano, de 9 a 7 meses, entre junho e dezembro, com níveis pluviométricos mais baixos, de 200 mm/ano. Como já mencionado, as temperaturas são geralmente elevadas. A temperatura média anual varia de 25° a 30°C e é aproximadamente constante em toda região. No período seco a temperatura do solo aumenta ainda mais, o que faz acelerar a evaporação das águas dos lagos e rios.

A Caatinga é caracterizada por possuir rios intermitentes ou temporários, que são aqueles que secam durante um período do ano. Comparados aos intermitentes, são poucos os rios perenes (que apresenta água corrente o ano todo) nesse bioma. Um exemplo de rio perene é o São Francisco. A fauna da Caatinga é rica em biodiversidade e a sua botânica é a menos conhecida em comparação com os demais biomas brasileiros. A Caatinga é um dos biomas mais devastados do Brasil, tendo a sua área destinada a atividades agropecuárias, que aumentam o desmatamento.

## **Cerrado**

O Cerrado ocupa 27% do território baiano é o segundo maior bioma da América Latina e do Brasil, além de ser reconhecido como a savana brasileira, apresentando uma grande biodiversidade e compreende uma área de elevado potencial aquífero. A vegetação do Cerrado apresenta tonalidades de verde, amarelo e tons amarronzados ocasionados pela descoloração causada pela incidência solar.

A vegetação do Cerrado se encontra em uma região do Estado da Bahia onde o clima predominantemente é o tropical sazonal, apresentando duas estações bem definidas: invernos secos e verões chuvosos. O período de seca tem início no mês de maio e termina no mês de setembro. Já o período chuvoso inicia-se em outubro e finaliza em abril. A precipitação média anual é de 1500 mm, variando de 750 a 2000mm. As temperaturas médias anuais variam do mínimo de 20° a 22°C até o máximo de 24° a 26°C.

No Cerrado encontram-se as nascentes dos principais rios brasileiros, constando, nove das doze bacias hidrográficas existentes no Brasil, de acordo com o IBGE. Além disso, o Cerrado está localizado numa região em que existem grandes aquíferos, como o Bambuí, na Bahia.

O Cerrado é considerado atualmente como o segundo bioma do Brasil mais ameaçado, ficando atrás somente da Mata Atlântica. As principais causas da devastação do Cerrado são o avanço das queimadas e o desmatamento para a utilização do solo na agropecuária. A área de ocupação original do Cerrado configura-se, hoje, como o principal local produtor de grãos do Brasil, com destaque para a soja, que é voltada, principalmente, para o mercado externo. Além da soja, os principais produtos agrícolas cultivados no Cerrado, são: a cana-de-açúcar, o milho e o algodão.

## **Mata Atlântica**

O bioma Mata Atlântica ocupa cerca de 19% do território baiano. Ele é formado por diferentes ecossistemas florestais e apresenta uma grande biodiversidade. O clima da Mata Atlântica, presente no estado da Bahia, é o tropical litorâneo úmido. Esse clima é caracterizado por temperaturas elevadas, altos índices pluviométricos e elevada umidade do ar, com escassez de períodos de estiagem.

A Mata Atlântica possui uma grande biodiversidade. Os seus solos são geralmente rasos e ácidos, extremamente úmidos e pobres. A pouca profundidade do solo somado aos altos níveis pluviométricos propiciam processos erosivos e deslizamentos nas partes mais altas. Como a colonização do Brasil teve início a partir do litoral, a Mata Atlântica foi a primeira a ser explorada com desmatamentos sucessivos a partir da extração de pau-brasil, e dos ciclos econômicos como o da cana-de-açúcar, café e ouro. Nos dias atuais, a mata Atlântica ainda sofre diversos impactos, como: o ambiental causado pela população que habita sua área; a agricultura e agropecuária; a exploração predatória de madeira e

espécies vegetais; a industrialização; a expansão urbana desordenada; o consumo excessivo; o lixo e a poluição.

Estas características gerais servem como suporte teórico para entender os padrões exibidos nas redes construídas a partir dos dados das variáveis climáticas.

# Referências

- [1] NICOSIA, V.; TANG, J.; MUSOLESI, M.; RUSSO, G.; MASCOLO, C.; LATORA, V. Components in time-varying graphs. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, v. 22, n. 2, p. 023101, 2012.
- [2] BOCCALETTI, S.; BIANCONI, G.; CRIADO, R.; DEL GENIO, C. I.; GÓMEZ-GARDENES, J.; ROMANCE, M.; SENDINA-NADAL, I.; WANG, Z.; ZANIN, M. The structure and dynamics of multilayer networks. *Physics Reports*, v. 544, n. 1, p. 1–122, 2014.
- [3] KIVELÄ, M.; ARENAS, A.; BARTHELEMY, M.; GLEESON, J. P.; MORENO, Y.; PORTER, M. A. Multilayer networks. *Journal of complex networks*, v. 2, n. 3, p. 203–271, 2014.
- [4] DE DOMENICO, M.; SOLÉ-RIBALTA, A.; COZZO, E.; KIVELÄ, M.; MORENO, Y.; PORTER, M. A.; GÓMEZ, S.; ARENAS, A. Mathematical formulation of multilayer networks. *Physical Review X*, v. 3, n. 4, p. 041022, 2013.
- [5] ROSÁRIO, R.; CARDOSO, P.; MUÑOZ, M.; MONTTOYA, P.; MIRANDA, J. Motif-synchronization: A new method for analysis of dynamic brain networks with eeg. *Physica A: Statistical Mechanics and its Applications*, v. 439, p. 7–19, 2015.
- [6] TEIXEIRA, G. M.; AGUIAR, M. D. S. F. D.; CARVALHO, C. F. D.; DANTAS, D. R.; CUNHA, M. D. V.; MORAIS, J. H. M. D.; PEREIRA, H. B. D. B.; MIRANDA, J. G. V. Complex semantic networks. *International Journal of Modern Physics C*, v. 21, n. 03, p. 333–347, 2010.
- [7] NEWMAN, M. E. J. *Networks: an introduction*. 1. ed. New York: OUP Oxford, 2010. v. volume único.
- [8] NEWMAN, M. E. The structure of scientific collaboration networks. *Proceedings of the national academy of sciences*, v. 98, n. 2, p. 404–409, 2001.
- [9] PIMM, S. L. *The balance of nature?: ecological issues in the conservation of species and communities*. 1. ed. Chicago: University of Chicago Press, 1991. v. volume único.
- [10] SZWARCFITER, J. L. *Grafos e algoritmos computacionais*. 2. ed. Rio de Janeiro: Editora Campus Ltda, 1984. v. volume único.
- [11] GERSTING, J. L. *Fundamentos matemáticos para a ciência da computação: um tratamento moderno de matemática discreta*. 3. ed. Rio de Janeiro: Livros Técnicos e Científicos, 2004. v. volume único.

- [12] NEWMAN, M. E. J. The structure and function of complex networks. *SIAM Review*, v. 45, n. 2, p. 167, 2003.
- [13] BOCCALETTI, S.; LATORA, V.; MORENO, Y.; CHAVEZ, M.; HWANG, D.-U. Complex networks: Structure and dynamics. *Physics Reports*, v. 424, n. 4-5, p. 175, 2006.
- [14] BARABÁSI, A.-L. *Linked (conectado): a nova ciência dos networks*. 1. ed. São Paulo: Leopardo, 2009. v. volume único.
- [15] BARABÁSI, A.-L.; ALBERT, R. Emergence of scaling in random networks. *science*, v. 286, n. 5439, p. 509–512, 1999.
- [16] HOLME, P.; SARAMÄKI, J. Temporal networks. *Physics reports*, v. 519, n. 3, p. 97–125, 2012.
- [17] KOSTAKOS, V. Temporal graphs. *Physica A: Statistical Mechanics and its Applications*, v. 388, n. 6, p. 1007–1023, 2009.
- [18] CASTEIGTS, A.; FLOCCHINI, P.; QUATTROCIOCCHI, W.; SANTORO, N. Time-varying graphs and dynamic networks. *International Journal of Parallel, Emergent and Distributed Systems*, v. 27, n. 5, p. 387, 2012.
- [19] WEHMUTH, K.; ZIVIANI, A.; FLEURY, E. A unifying model for representing time-varying graphs. In: . c2015. p. 1–10.
- [20] SOUSA, R.; LULA-ROCHA, V.; TOUTAIN, T.; ROSÁRIO, R.; CAMBUI, E.; MIRANDA, J. Preferential interaction networks: A dynamic model for brain synchronization networks. *Physica A: Statistical Mechanics and its Applications*, v. 554, p. 124259, 2020.
- [21] DO VALE CUNHA, M.; SANTOS, C. C. R.; MORET, M. A.; DE BARROS PEREIRA, H. B. Shannon entropy in time-varying semantic networks of titles of scientific paper. *Applied Network Science*, v. 5, n. 1, p. 1–17, 2020.
- [22] SANTORO, N.; QUATTROCIOCCHI, W.; FLOCCHINI, P.; CASTEIGTS, A.; AMBLARD, F. Time-varying graphs and social network analysis: Temporal indicators and metrics. *CoRR*, v. abs/1102.0629, 2011.
- [23] SABA, H.; VALE, V. C.; MORET, M. A.; MIRANDA, J. G. V. Spatio-temporal correlation networks of dengue in the state of bahia. *BMC Public Health*, v. 14, n. 1, p. 1085, 2014.
- [24] ROSÁRIO, R. S. *Sincronização por motifs uma proposta para análise de redes cerebrais*. 2013. p. 73. Dissertação (Mestrado em Física) - Universidade Federal da Bahia, Salvador, 2013.

- [25] BARRETO, C. S. F. *Uso dos grafos variantes no tempo no estudo do efeito da estimulação transcraniana de corrente contínua em redes funcionais corticais*. 2015. p. 55. Dissertação (Mestrado em Física) - Universidade Federal da Bahia, Salvador, 2015.
- [26] BERTONI, V.; SAURIN, T.; FOGLIATTO, F.; FALEGNAMI, A.; PATRIARCA, R. Monitor, anticipate, respond, and learn: Developing and interpreting a multilayer social network of resilience abilities. *Safety Science*, v. 136, n. 105148, 2021.
- [27] SMITH-AGUILAR, S. E.; AURELI, F.; BUSIA, L.; SCHAFFNER, C.; RAMOS-FERNÁNDEZ, G. Using multiplex networks to capture the multidimensional nature of social structure. *Primates*, v. 60, n. 3, p. 277–295, 2019.
- [28] VAROL, O.; MENCZER, F. Connecting dream networks across cultures. In: . c2014. p. 1267–1272.
- [29] UZZI, B.; MUKHERJEE, S.; STRINGER, M.; JONES, B. Atypical combinations and scientific impact. *Science*, Washington, v. 342, n. 6157, p. 468–472, 2013.
- [30] NANUMYAN, V.; GOTE, C.; SCHWEITZER, F. Multilayer network approach to modeling authorship influence on citation dynamics in physics journals. *Physical Review E*, v. 102, n. 3, p. 032303, 2020.
- [31] HORVÁT, E.-Á.; ZWEIG, K. A. A fixed degree sequence model for the one-mode projection of multiplex bipartite graphs. *Social Network Analysis and Mining*, v. 3, n. 4, p. 1209–1224, 2013.
- [32] HORVÁT, E.-A.; ZWEIG, K. A. One-mode projection of multiplex bipartite graphs. In: . c2012. p. 599–606.
- [33] KAZIENKO, P.; MUSIAL, K.; KAJDANOWICZ, T. Multidimensional social network in the social recommender system. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, v. 41, n. 4, p. 746–759, 2011.
- [34] NAZIR, A.; RAZA, S.; CHUAH, C.-N. Unveiling facebook: a measurement study of social network based applications. In: . c2008. p. 43–56.
- [35] TANG, L.; WANG, X.; LIU, H. Community detection via heterogeneous interaction analysis. *Data mining and knowledge discovery*, v. 25, n. 1, p. 1–33, 2012.
- [36] ALETA, A.; MELONI, S.; MORENO, Y. A multilayer perspective for the analysis of urban transportation systems. *Scientific reports*, v. 7, n. 1, p. 1–9, 2017.
- [37] MORRIS, R. G.; BARTHELEMY, M. Transport on coupled spatial networks. *Physical review letters*, v. 109, n. 12, p. 128703, 2012.

- [38] GRANELL, C.; GÓMEZ, S.; ARENAS, A. Dynamical interplay between awareness and epidemic spreading in multiplex networks. *Physical review letters*, v. 111, n. 12, p. 128701, 2013.
- [39] SAUMELL-MENDIOLA, A.; SERRANO, M. Á.; BOGUNÁ, M. Epidemic spreading on interconnected networks. *Physical Review E*, v. 86, n. 2, p. 026106, 2012.
- [40] GÓMEZ-GARDENES, J.; REINARES, I.; ARENAS, A.; FLORÍA, L. M. Evolution of cooperation in multiplex networks. *Scientific reports*, v. 2, p. 620, 2012.
- [41] MUCHA, P. J.; RICHARDSON, T.; MACON, K.; PORTER, M. A.; ONNELA, J.-P. Community structure in time-dependent, multiscale, and multiplex networks. *Science*, Washington, v. 328, n. 5980, p. 876–878, 2010.
- [42] BIANCONI, G. Statistical mechanics of multiplex networks: Entropy and overlap. *Physical Review E*, v. 87, n. 6, p. 062806, 2013.
- [43] BARIGOZZI, M.; FAGIOLO, G.; GARLASCHELLI, D. Multinetwork of international trade: A commodity-specific analysis. *Physical Review E*, v. 81, n. 4, p. 046104, 2010.
- [44] BARIGOZZI, M.; FAGIOLO, G.; MANGIONI, G. Identifying the community structure of the international-trade multi-network. *Physica A: statistical mechanics and its applications*, v. 390, n. 11, p. 2051–2066, 2011.
- [45] BARRETT, L.; HENZI, S. P.; LUSSEAU, D. Taking sociality seriously: the structure of multi-dimensional social networks as a source of information for individuals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, v. 367, n. 1599, p. 2108–2118, 2012.
- [46] NELSON, D. L.; MCEVOY, L., C.; SCHREIBER, T. A. The university of south florida word association, rhyme, and word fragment norms, 1998.
- [47] GUTIÉRREZ, R.; SENDINA-NADAL, I.; ZANIN, M.; PAPO, D.; BOCCALETTI, S. Targeting the dynamics of complex networks. *Scientific reports*, v. 2, n. 1, p. 1–5, 2012.
- [48] COSTA, L. D. F.; OLIVEIRA JR, O. N.; TRAVIESO, G.; RODRIGUES, F. A.; VILLAS BOAS, P. R.; ANTIQUEIRA, L.; VIANA, M. P.; CORREA ROCHA, L. E. Analyzing and modeling real-world phenomena with complex networks: a survey of applications. *Advances in Physics*, London, v. 60, n. 3, p. 329–412, 2011.
- [49] MEIDAN, D.; SCHULMANN, N.; COHEN, R.; HABER, S.; YANIV, E.; SARID, R.; BARZEL, B. Alternating quarantine for sustainable epidemic mitigation. *Nature communications*, v. 12, n. 1, p. 1–12, 2021.



- [50] BAGLER, G. Analysis of the airport network of india as a complex weighted network. *Physica A: Statistical Mechanics and its Applications*, v. 387, n. 12, p. 2972, 2008.
- [51] LEE, K.-M.; MIN, B.; GOH, K.-I. Towards real-world complexity: an introduction to multiplex networks. *The European Physical Journal B*, Les Ulis, v. 88, n. 2, p. 1–20, 2015.
- [52] DEL GENIO, C. I.; GÓMEZ-GARDEÑES, J.; BONAMASSA, I.; BOCCALETTI, S. Synchronization in networks with multiple interaction layers. *Science advances*, v. 2, n. 11, p. e1601679, 2016.
- [53] RAKSHIT, S.; MAJHI, S.; BERA, B. K.; SINHA, S.; GHOSH, D. Time-varying multiplex network: Intralayer and interlayer synchronization. *Physical Review E*, v. 96, n. 6, p. 062308, 2017.
- [54] SZELL, M.; LAMBIOTTE, R.; THURNER, S. Multirelational organization of large-scale social networks in an online world. *Proceedings of the National Academy of Sciences*, v. 107, n. 31, p. 13636–13641, 2010.
- [55] PARSHANI, R.; ROZENBLAT, C.; IETRI, D.; DUCRUET, C.; HAVLIN, S. Inter-similarity between coupled networks. *EPL (Europhysics Letters)*, v. 92, n. 6, p. 68002, 2011.
- [56] HALU, A.; MUKHERJEE, S.; BIANCONI, G. Emergence of overlap in ensembles of spatial multiplexes and statistical mechanics of spatial interacting network ensembles. *Physical Review E*, v. 89, n. 1, p. 012806, 2014.
- [57] LEVIS, D.; PAGONABARRAGA, I.; DÍAZ-GUILERA, A. Synchronization in dynamical networks of locally coupled self-propelled oscillators. *Physical Review X*, v. 7, n. 1, p. 011028, 2017.
- [58] RAKKIYAPPAN, R.; SAKTHIVEL, N. Pinning sampled-data control for synchronization of complex networks with probabilistic time-varying delays using quadratic convex approach. *Neurocomputing*, v. 162, p. 26–40, 2015.
- [59] DÖRFLER, F.; BULLO, F. Synchronization in complex networks of phase oscillators: A survey. *Automatica*, v. 50, n. 6, p. 1539–1564, 2014.
- [60] JALAN, S.; SINGH, A.; ACHARYYA, S.; KURTHS, J. Impact of a leader on cluster synchronization. *Physical Review E*, v. 91, n. 2, p. 022901, 2015.
- [61] ROSSLER, O. An equation for continuous chaos. *Physics Letters A*, Amsterdam, v. 57, n. 5, p. 397 – 398, 1976.
- [62] Instituto brasileiro de geografia e estatística. <https://www.ibge.gov.br/cidades-e-estados/ba.html>. acessado em 28/07/2022.

- [63] JUNIOR, L. A. F. Guia técnico para a recuperação de vegetação em imóveis rurais no estado da bahia, 2017.
- [64] DUTRA, G. F.; FRANÇOIS-TIMMERS, J.; MESQUITA, C. A. B.; BEDÊ, L.; PINHEIRO, T.; PINTO, L. Biodiversidade e desenvolvimento na bahia. *Bahia Análise & Dados, Salvador*, v. 22, n. 3, p. 485–502, 2012.
- [65] ALLEN, R. G.; PEREIRA, L. S.; RAES, D.; SMITH, M. et al. Evapotranspiración del cultivo: guías para la determinación de los requerimientos de agua de los cultivos. *Roma: FAO*, v. 298, n. 0, 2006.
- [66] CRUZ, J. T. D. Avaliação de métodos de estimativa de evapotranspiração para Brasília-df. Brasília-DF, p. 45, 2016.
- [67] ALLEN, R. G.; PEREIRA, L. S.; RAES, D.; SMITH, M. et al. Crop evapotranspiration-guidelines for computing crop water requirements-fao irrigation and drainage paper 56. *Fao, Rome*, v. 300, n. 9, p. D05109, 1998.
- [68] ROUSE JR, J.; HAAS, R. H.; DEERING, D.; SCHELL, J.; HARLAN, J. C. Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation. Technical report, 1974.
- [69] HATFIELD, J. L.; GITELSON, A. A.; SCHEPERS, J. S.; WALTHALL, C. L. Application of spectral remote sensing for agronomic decisions. *Agronomy Journal*, v. 100, n. S3, p. S–117–S–131, 2008.
- [70] Google earth engine. <https://developers.google.com/earth-engine/edu>. acessado em 21/01/2022.
- [71] GORELICK, N.; HANCHER, M.; DIXON, M.; ILYUSHCHENKO, S.; THAU, D.; MOORE, R. Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote sensing of Environment*, v. 202, p. 18–27, 2017.
- [72] RUNNING, S. W.; MU, Q.; ZHAO, M.; MORENO, A. Modis global terrestrial evapotranspiration (et) product (mod16a2/a3 and year-end gap-filled mod16a2gf/a3gf) nasa earth observing system modis land algorithm (for collection 6). *National Aeronautics and Space Administration, Washington, DC, USA [data set]*, <https://doi.org/10.5067/MODIS/MOD16A2>, v. 6, 2019.
- [73] Earth engine data catalog. [https://developers.google.com/earth-engine/datasets/catalog/MODIS\\_006\\_MOD16A2#description](https://developers.google.com/earth-engine/datasets/catalog/MODIS_006_MOD16A2#description). acessado em 23/10/2021.
- [74] HUETE, A.; JUSTICE, C.; VAN LEEUWEN, W. Modis vegetation index (mod13). *Algorithm theoretical basis document*, v. 3, n. 213, p. 295–309, 1999.

- [75] Earth engine data catalog. [https://developers.google.com/earth-engine/datasets/catalog/MODIS\\_006\\_MOD13A1](https://developers.google.com/earth-engine/datasets/catalog/MODIS_006_MOD13A1). acessado em 23/10/2021.
- [76] FUNK, C.; PETERSON, P.; LANDSFELD, M.; PEDREROS, D.; VERDIN, J.; SHUKLA, S.; HUSAK, G.; ROWLAND, J.; HARRISON, L.; HOELL, A. et al. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific data*, v. 2, n. 1, p. 1–21, 2015.
- [77] Earth engine data catalog. [https://developers.google.com/earth-engine/datasets/catalog/UCSB-CHG\\_CHIRPS\\_DAILY#description](https://developers.google.com/earth-engine/datasets/catalog/UCSB-CHG_CHIRPS_DAILY#description). acessado em 23/10/2021.
- [78] WAN, Z. Modis land-surface temperature algorithm basis document (1st atbd): version 3.3. <http://modis.gsfc.nasa.gov/data/atbd>, 1999.
- [79] Earth engine data catalog. [https://developers.google.com/earth-engine/datasets/catalog/MODIS\\_006\\_MOD11A2#description](https://developers.google.com/earth-engine/datasets/catalog/MODIS_006_MOD11A2#description). acessado em 23/10/2021.
- [80] BAIANO, S. Caracterização da variabilidade espacial e temporal da precipitação pluviométrica do centro. *Revista Brasileira de Geografia Física*, v. 10, n. 4, p. 1143–1151, 2017.
- [81] DOURADO, C. D. S.; DE ÁVILA, A. M.; OLIVEIRA, S. D. M. Regionalização da precipitação no estado da bahia por meio de técnicas de mineração de dados. In: . c2012.
- [82] BLONDEL, V. D.; GUILLAUME, J.-L.; LAMBIOTTE, R.; LEFEBVRE, E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, v. 2008, n. 10, p. P10008, oct 2008.
- [83] BUAINAIN, A.; FAVARETO, A.; CONTINI, E.; CHAVES, F.; HENZ, G.; GARCIA, J.; DAMIANI, O.; VIEIRA JUNIOR, P.; GRUNDLING, R.; NOGUEIRA, V. D. C. Desafios para agricultura nos biomas brasileiros. *Área de Informação da Sede-Livro científico (ALICE)*, 2020.