

**UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS**

João Vitor Bevilacqua de Souza Merenda

**Reconhecimento de padrões em redes complexas usando
caminhadas determinísticas do turista**

São Carlos

2023

João Vitor Bevilacqua de Souza Merenda

**Reconhecimento de padrões em redes complexas usando
caminhadas determinísticas do turista**

Dissertação apresentada ao Programa de Pós-Graduação em Física do Instituto de Física de São Carlos da Universidade de São Paulo, para obtenção do título de Mestre em Ciências.

Área de concentração: Física Computacional

Orientador: Prof. Dr. Odemir Martinez Bruno

Versão corrigida

(Versão original disponível na Unidade que aloja o Programa)

São Carlos

2023

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Merenda, João Vitor Bevilacqua de Souza
Reconhecimento de padrões em redes complexas usando
caminhadas determinísticas do turista / João Vitor
Bevilacqua de Souza Merenda; orientador Odemir Martinez
Bruno - versão corrigida -- São Carlos, 2023.
101 p.

Dissertação (Mestrado - Programa de Pós-Graduação em
Física Computacional) -- Instituto de Física de São Carlos,
Universidade de São Paulo, 2023.

1. Redes Complexas. 2. Reconhecimento de Padrões. 3.
Autômatos. 4. Caminhada determinística do turista. I.
Martinez Bruno, Odemir, orient. II. Título.

*Este trabalho é dedicado à minha irmã Izadora Bevilacqua de Souza Merenda
que faleceu em 2012 vítima de Leucemia.*

AGRADECIMENTOS

Agradeço primeiramente a Deus por me dar a vida, saúde e capacidade.

Agradeço aos meus pais pelo incansável apoio durante o desenvolvimento deste trabalho.

À minha família por todo o incentivo e pelas palavras de carinho.

Ao Prof. Dr. Odemir Martinez Bruno pela atenção e apoio durante o processo de orientação.

Ao Instituto de Física de São Carlos da Universidade de São Paulo pela oportunidade de realização do curso de mestrado.

Por fim, agradeço à Universidade de São Paulo por todo o conhecimento que adquiri durante a graduação e a pós-graduação.

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

“I think the next century will be the century of complexity.”
Stephen Hawking

RESUMO

MERENDA, J. V. B. S. **Reconhecimento de padrões em redes complexas usando caminhadas determinísticas do turista.** 2023. 101p. Dissertação (Mestrado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2023.

A ciência de redes tem tornado-se cada vez mais presente, tanto em meios acadêmicos quanto na indústria e no cotidiano. Diversos sistemas reais, considerados complexos, podem ser reduzidos a uma rede, onde o formalismo exibe um número pequeno de variáveis. Dentre os sistemas que podem ser representados por redes estão: as redes sociais, redes de reações químicas, teias alimentares e redes neurais. Dado o elevado número de aplicações no mundo real, tornou-se imprescindível reconhecer padrões, encontrar variáveis determinantes que definem a estrutura da rede. Nessa dissertação abordaremos dois métodos para o reconhecimento de padrão em redes. O primeiro foi desenvolvido para estudar redes de pequeno-mundo e o segundo para analisar tanto redes sintéticas quanto redes reais. O primeiro método mostrou bons resultados em mostrar a transição de rede regular para rede aleatória em grafos de Watts-Strogatz. O segundo método, chamado de caminhada determinística do turista com bifurcações, mostrou bom desempenho na classificação de redes sintéticas e de redes reais.

Palavras-chave: Redes complexas. Reconhecimento de padrões. Autômatos. Caminhada determinística do turista.

ABSTRACT

MERENDA, J. V. B. S. **Pattern recognition in complex networks using deterministic tourist walks**. 2023. 101p. Dissertation (Master in Science) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2023.

Network science has become increasingly present in academic circles, industry, and everyday life. Many real-world systems, considered complex systems, can be reduced to a network since this formalism exhibits a short number of variables. Some systems, such as social networks, chemical reaction networks, food webs, and neural networks, can be represented by networks. Therefore, it has become fundamental to recognize patterns and find a variable set that defines the network structure. In this dissertation, we will discuss two methods for pattern recognition in networks. The first method was developed to study small-world networks, and the second algorithm was made to analyze both synthetic and real-world networks. The first method presented well results in verifying the transition from regular network to random network in Watts-Strogatz graphs. The second method, called deterministic tourist walk with bifurcations, got a good performance in the synthetic and real-world network classification.

Keywords: Complex networks. Pattern recognition. Automata. Deterministic tourist walk.

LISTA DE FIGURAS

Figura 1 – O problema das sete pontes de <i>Königsberg</i> . O problema pode ser convertido em um grafo, como mostra a figura da direita.	24
Figura 2 – Tipos de grafos. (a) Grafo não dirigido. (b) Grafo dirigido (digrafo). (c) Grafo com pesos.	27
Figura 3 – Medidas de centralidade.	34
Figura 4 – Modelo de Watts-Strogatz. (a) Rede regular. (b) Rede de pequeno-mundo. (c) Rede aleatória. (d) Curvas de média do comprimento dos menores caminhos (azul) e coeficiente de aglomeração global (vermelho).	38
Figura 5 – Distribuição de grau de redes do mundo real.	41
Figura 6 – Três modelos de redes geométricas. (a) Rede de Voronoi, incluindo as células de Voronoi. (b) Rede geométrica aleatória ($R = 0.25$). (c) Rede de Waxman ($\alpha = 0.1$).	42
Figura 7 – Quatro modelos de redes e suas respectivas distribuições de grau.	43
Figura 8 – Exemplo de trajetória de uma caminhada determinística do turista em: (a) Redes. (b) Imagens. As setas em azul indicam a parte transiente da trajetória e as setas em vermelho representam o atrator.	49
Figura 9 – (a) Imagem original. (b) Imagem produzida através de uma rede complexa modelada pelo algoritmo DTW. Os tons de cinza equivalem ao valor de grau do vértice correspondente ao pixel. Para uma maior resolução, os valores de grau foram convertidos numa escala de 0 à 255 usando uma transformação logarítmica.	50
Figura 10 – Caminhada determinística do turista com bifurcações (DTWB)	56
Figura 11 – Grau de bifurcação	57
Figura 12 – Quantidade de trajetórias geradas, partindo de um único nó, a cada iteração. Rede: Watts-Strogatz, com 500 nós e grau médio igual a seis. O turista possui memória $\mu = 2$	66
Figura 13 – Os quatro principais modelos de redes (linha de cima) e suas respectivas distribuições de grau de bifurcação B (linha de baixo). (a) Rede BA. (b) Rede ER. (c) Rede de Waxman. (d) Rede WS. Regra de caminhada: B-R1 e memória $\mu = 1$. Cada rede possui 500 nós e grau topológico médio igual a seis.	68
Figura 14 – Os quatro principais modelos de redes (linha de cima) e suas respectivas distribuições de grau de bifurcação B (linha de baixo). (a) Rede BA. (b) Rede ER. (c) Rede de Waxman. (d) Rede WS. Regra de caminhada: B-R2 e memória $\mu = 1$. Cada rede possui 500 nós e grau topológico médio igual a seis.	69

Figura 15 – Os quatro principais modelos de redes (linha de cima) e suas respectivas distribuições de grau de bifurcação B (linha de baixo). (a) Rede BA. (b) Rede ER. (c) Rede de Waxman. (d) Rede WS. Regra de caminhada: B-R3 e memória $\mu = 1$. Cada rede possui 500 nós e grau topológico médio igual a seis.	71
Figura 16 – Os quatro principais modelos de redes (linha de cima) e suas respectivas distribuições de grau de bifurcação B (linha de baixo). (a) Rede BA. (b) Rede ER. (c) Rede de Waxman. (d) Rede WS. Regra de caminhada: B-R4 e memória $\mu = 1$. Cada rede possui 500 nós e grau topológico médio igual a seis.	71
Figura 17 – Classificação através do método PCA usando os vetores de característica produzidos pelo método DTWB para as regras B-R1 (linha de cima) e B-R2 (linha de baixo). Os eixos são as componentes principais.	81
Figura 18 – Matrizes de confusão da classificação das redes do <i>dataset</i> 1 e do <i>dataset</i> 2 para as regras B-R1 (linha de cima) e B-R2 (linha de baixo).	82
Figura 19 – <i>Scatter plot</i> dos vetores de característica (Eq. 3.6) para redes de Watts-Strogatz com probabilidade variando entre 10^{-4} (vermelho) e 10^0 (azul). 87	87
Figura 20 – Valores de comprimento da trajetória do turista (l), normalizado pelo comprimento da trajetória do turista para uma rede aleatória equivalente (lr). Foram usados quatro valores de memória: $\mu = 1, 2, 3, 4$. A linha tracejada em magenta mostra a evolução do coeficiente de aglomeração no mesmo intervalo de probabilidade de religação. Redes com $N = 500$ e $\langle k \rangle = 10$	87
Figura 21 – Valores de χ (eixo esquerdo) e ω (eixo direito) para redes de Watts-Strogatz para $p \in [10^{-4}, 10^0]$. As linhas sólidas em vermelho, verde e azul são os valores de χ para memória $\mu = 2, \mu = 3$ e $\mu = 4$, respectivamente. A linha tracejada em ciano são os valores do índice ω . Todas as redes possuem 500 nós.	88
Figura 22 – Comparação de tempo de execução (em segundos) entre a métrica χ e a métrica ω . (a) Comparação de tempo de execução para redes de mesmo tamanho n e mesmo grau médio $\langle K \rangle$, mas com probabilidade de religação variando. (b) Comparação de tempo de execução para redes com mesma probabilidade de religação $p = 0.25$, mas com tamanhos (n) diferentes.	90

LISTA DE TABELAS

Tabela 1 – Média e desvio padrão do grau de bifurcação B . Regra: B-R1	67
Tabela 2 – Média e desvio padrão da razão β . Regra: B-R1.	67
Tabela 3 – Média e desvio padrão do grau de bifurcação B . Regra: B-R2.	69
Tabela 4 – Média e desvio padrão da razão β . Regra: B-R2.	70
Tabela 5 – Média e desvio padrão do grau de bifurcação B . Regra: B-R3.	70
Tabela 6 – Média e desvio padrão da razão β . Regra: B-R3.	70
Tabela 7 – Média e desvio padrão do grau de bifurcação B . Regra: B-R4.	70
Tabela 8 – Média e desvio padrão da razão β em redes sintéticas. Regra B-R4 e memórias $\mu = 1, \mu = 2, \mu = 3$ e $\mu = 4$	72
Tabela 9 – Média e desvio padrão dos transientes. Regra: B-R1.	72
Tabela 10 – Média e o desvio padrão dos atratores. Regra: B-R1.	72
Tabela 11 – Média e o desvio padrão dos transientes. Regra: B-R2.	73
Tabela 12 – Média e o desvio padrão dos atratores. Regra: B-R2.	73
Tabela 13 – Média e o desvio padrão dos transientes. Regra: B-R3.	73
Tabela 14 – Média e o desvio padrão dos atratores. Regra: B-R3.	74
Tabela 15 – Média e o desvio padrão dos transientes. Regra: B-R4.	74
Tabela 16 – Média e o desvio padrão dos atratores. Regra: B-R4.	74
Tabela 17 – Desempenho do vetor de característica gerado através dos histogramas. $\mu = 1$. Redes do <i>Dataset 1</i>	77
Tabela 18 – Desempenho individual (%) das medidas utilizadas nos vetores de característica. $\mu = 1$. <i>Dataset</i> : Redes sintéticas.	77
Tabela 19 – Desempenho (%) das combinações de medidas no vetor de característica. $\mu = 1$	78
Tabela 20 – Valores de acurácia dos métodos utilizados como comparação ao método DTWB. o Método DTW usado nos experimentos utiliza quatro valores de memória combinados ($\vec{\mu} = [1, 2, 3, 4]$) para cada regra de caminhada.	79
Tabela 21 – Valores de acurácia do método DTWB para valores únicos de memória e regra B-R1.	80
Tabela 22 – Valores de acurácia do método DTWB para valores únicos de memória e regra B-R2.	80
Tabela 23 – Valores de acurácia do método DTWB para valores únicos de memória e regra B-R3.	82
Tabela 24 – Valores de acurácia do método DTWB para valores únicos de memória e regra B-R4.	83
Tabela 25 – Valores de acurácia do método DTWB combinando valores de memória para a regra B-R1.	83

Tabela 26 – Valores de acurácia do método DTWB combinando valores de memória para a regra B-R2.	84
Tabela 27 – Valores de acurácia do método DTWB combinando valores de memória para a regra B-R3.	84
Tabela 28 – Valores de acurácia do método DTWB combinando valores de memória para a regra B-R4.	85
Tabela 29 – Valores de acurácia do método DTWB combinando as regras de caminhada. Em cada caso também há combinação de memória: $\vec{\mu} = [1, 2, 3, 4]$. Comb1 = B-R1 \cup B-R2, Comb2 = B-R3 \cup B-R4 e Comb3 = B-R1 \cup B-R2 \cup B-R3 \cup B-R4.	85
Tabela 30 – Valores de acurácia do método DTWB combinando as regras B-R1 e B-R2 e para seis valores de memória combinados $\vec{\mu} = [1, 2, 3, 4, 5, 6]$. . .	89

LISTA DE ABREVIATURAS E SIGLAS

BA	Barabasi-Albert
B-R1	<i>Bifurcation-Rule 1</i>
B-R2	<i>Bifurcation-Rule 2</i>
DTW	<i>Deterministic Tourist Walk</i> (Caminhada determinística do turista)
DTWB	<i>Deterministic Tourist Walk with Bifurcations</i> (Caminhada determinística do turista com bifurcações)
ER	Erdos-Renyi
IFSC	Instituto de Física de São Carlos
LDA	<i>Linear Discriminant Analysis</i>
LHC	<i>Large Hadron Collider</i>
LOOCV	<i>Leave-One-Out Cross-Validation</i>
Max	Máximo
Min	Mínimo
ML	<i>Machine Learning</i>
NWS	Newman-Watts-Strogatz
PCA	<i>Principal Component Analysis</i>
Gráfico Q-Q	Gráfico Quantil-Quantil
RGG	<i>Random geometric Graph</i>
SW	<i>Small-World</i>
SW-R1	<i>Small-World-Rule 1</i>
SW-R2	<i>Small-World-Rule 2</i>
T-R1	<i>Texture-Rule 1</i>
T-R2	<i>Texture-Rule 2</i>
USP	Universidade de São Paulo

Wax

Waxman

WS

Watts-Strogatz

LISTA DE SÍMBOLOS

α	Letra grega Alfa
β	Letra grega Beta
γ	Letra grega Gama
λ	Letra grega Lambda
μ	Letra grega Mi
π	Letra grega Pi
ρ	Letra grega Rô
σ	Letra grega minúscula Sigma
Σ	Letra grega maiúscula Sigma
τ	Letra grega Tau
ϕ	Letra grega Fi
χ	Letra grega Qui
ψ	Letra grega Psi
ω	Letra grega minúscula ômega
Ω	Letra grega maiúscula Ômega
\in	Pertence
\cup	União
\leq	Menor ou igual
\geq	Maior ou igual
\approx	Aproximadamente

SUMÁRIO

1	INTRODUÇÃO	23
1.1	Motivações para o reconhecimento de padrões em redes	25
1.2	Objetivo	26
2	CIÊNCIA DE REDES	27
2.1	Propriedades e medidas de redes	27
2.1.1	Medidas de centralidade	31
2.2	Modelos de redes	33
2.2.1	Redes regulares	34
2.2.2	Redes aleatórias	35
2.2.3	Redes de pequeno-mundo	36
2.2.4	Redes livre de escala	39
2.2.5	Redes geométricas	41
2.3	Métodos de reconhecimento de padrão em redes	42
2.3.1	Métodos baseados nas medidas e estruturas da rede	42
2.3.2	Métodos baseados em autômatos	43
3	A CAMINHADA DETERMINÍSTICA DO TURISTA	47
3.1	O método DTW aplicado à análise de texturas	49
3.2	O método DTW aplicado à redes	51
3.3	O método DTW aplicado à redes de pequeno-mundo (DTW-SW)	51
4	A CAMINHADA DETERMINÍSTICA DO TURISTA COM BIFURCAÇÕES	55
4.1	O vetor de característica	58
4.2	O método DTWB é determinístico?	59
5	MÉTODO EXPERIMENTAL	61
5.1	Datasets	61
5.2	Métodos de classificação e validação	62
5.3	Métodos de comparação	62
5.4	Comparação de tempo computacional	63
6	RESULTADOS E DISCUSSÕES	65
6.1	Complexidade computacional	65
6.2	Análise do método DTWB	66
6.2.1	Grau de bifurcação (B) e densidade de bifurcação (β)	66

6.2.2	Transiente e atrator	72
6.3	O vetor de característica: Outras medidas para o método DTWB .	75
6.4	Reconhecimento de padrões	79
6.5	Métricas de Pequeno-mundo baseadas na caminhada do turista . .	86
7	CONSIDERAÇÕES FINAIS	91
	REFERÊNCIAS	93
	APÊNDICES	97
	APÊNDICE A – MÉTODOS DE CLASSIFICAÇÃO E VALIDAÇÃO:	
	LDA, PCA E LOOCV	99
A.1	Linear Discriminant Analysis	99
A.2	Principal Component Analysis	100
A.3	Método Leave-One-Out cross-validation	101

1 INTRODUÇÃO

O reconhecimento de padrões é uma área cada vez mais importante no mundo moderno e que une diversos campos do conhecimento, tais como estatística, computação e teoria da informação. O principal objetivo do reconhecimento de padrões é a classificação de objetos em classes ou categorias (1), onde os objetos podem ser imagens, sinais em forma de ondas (*e.g.*, ondas de rádio, som) ou conjuntos de dados obtidos a partir de um experimento em física, por exemplo, o imenso conjunto de dados extraído dos experimentos no LHC. Desta forma, o reconhecimento de padrões está presente em muitos campos, tais como visão computacional, física experimental, astrofísica, engenharia, astrobiologia, bem como em ciência de dados e todas as áreas que dependem dela, dentre elas a economia, publicidade e ciências sociais. O reconhecimento de padrões também é usado em sistemas de inteligência artificial (2), naqueles que necessitam reconhecer imagens (visão computacional) e naqueles que lidam com uma tomada autônoma de decisão, como os carros autônomos, onde as classes são as decisões a serem tomadas (*e.g.* parar, acelerar, virar à direita, virar à esquerda).

Embora exista de forma teórica desde a década de 1960, essa área ganhou maior ênfase para aplicações a partir do advento da computação e das técnicas de aprendizado de máquinas ou *machine learning*, em inglês. Para converter um objeto ou sinal do mundo real em dados que máquinas possam entender, precisamos encontrar um conjunto de características, ou atributos, que descreva tal objeto. O conjunto de atributos forma um vetor de atributos. Por exemplo, suponha que tenhamos N frutas que precisam ser classificadas em C classes, *e.g.*, laranjas, maçãs e bananas. Podemos pensar na cor como um atributo importante. Também podemos incluir o gosto, o cheiro e a forma. Teremos, portanto, um vetor com m características para cada fruta. Por fim, esse conjunto de N vetores será classificado por um método de aprendizado de máquinas. Dentre esses métodos estão os mais tradicionais, como o *Linear discriminant analysis* (LDA), *Principal component analysis* (PCA), *Naive-Bayes* e os recentes métodos baseados em redes neurais e *Deep Learning*.

Outra área que evoluiu com o advento dos computadores foi a ciência de redes. A ciência de redes é a união da teoria dos grafos, estatística e computação. Dentre seus objetivos está o estudo das redes complexas, incluindo sistemas do mundo real, tais como a internet, redes sociais, redes de infraestrutura, redes neurais, entre outros. Uma rede é representada por grafos, possuindo um conjunto de vértices ou nós e um conjunto de conexões, ou arestas, entre esses vértices. O primeiro uso de redes para resolver problemas do mundo real data de 1736, quando o matemático Leonhard Euler utilizou a teoria dos grafos para tentar resolver o problema das sete pontes de *Königsberg*. (3) A cidade de

Königsberg, na antiga Prússia (hoje chamada de Kaliningrado, pertencente a Rússia) é situada sobre o rio *Pregel* e, no meio do fluxo do rio, há duas ilhas. Sete pontes conectam as quatro partes da cidade como mostra a Figura 1. O problema consistia em encontrar um caminho que passa por todas as sete pontes uma única vez. Utilizando um grafo, Euler representou as quatro massas de terra como vértices e as pontes como aresta (ver Figura 1). Euler mostrou que um caminho que passa por todas as pontes uma única vez, também chamado de caminho euleriano, somente seria possível se houvesse exatamente zero ou dois vértices com um número ímpar de arestas conectadas a eles e os demais com um número par de arestas. Em outras palavras, Euler demonstrou que a possibilidade de solução do problema depende do grau dos vértices. Dessa forma, Euler mostrou que não é possível um caminho euleriano sobre as sete pontes de *Königsberg*, uma vez que todos os vértices possuem um número ímpar de arestas, como mostra a Figura 1.

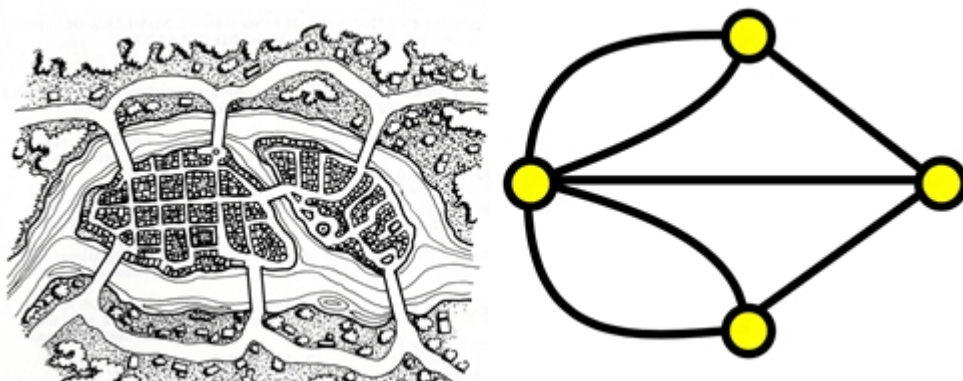


Figura 1 – O problema das sete pontes de *Königsberg*. O problema pode ser convertido em um grafo, como mostra a figura da direita.

Fonte: MATHEMATICAL ASSOCIATION OF AMERICA (4)

Dois séculos mais tarde, em 1951, Anatol Rapoport e Ray Solomonoff tornaram-se pioneiros em estudar interações sociais através de redes, chamadas hoje de redes sociais (5). O trabalho de Rapoport e Solomonoff mostrou como a conectividade de uma rede facilita a disseminação de informação, epidemias, bem como a difusão de inovação. Nesse trabalho, os autores introduziram as redes aleatórias, as quais seriam alguns anos mais tarde formalizadas por Paul Erdos, Alfred Renyi (6) e Edgar Gilbert. (7) O estudo das redes sociais e de diversos sistemas ocorridos na segunda metade do século XX mostraram o enorme potencial de aplicação das redes complexas no estudo de sistemas reais. Com a evolução do computador, a ciência de redes tornou-se cada vez mais presente no meio acadêmico. Vários modelos surgiram nas últimas três décadas e diversos algoritmos foram desenvolvidos para extrair características de redes.

Um dos algoritmos mais recentes para extrair características de redes e também de imagens é a caminhada determinística do turista, ou *Deterministic tourist walk* (DTW),

em inglês. (8) Nesse algoritmo, um autômato, também chamado de turista ou caminhante, caminha de vértice a vértice seguindo regras determinísticas de caminhada, diferenciando-se, portanto, das caminhadas aleatórias, onde a direção de caminhada é tomada através de uma probabilidade de transição.

1.1 Motivações para o reconhecimento de padrões em redes

Existem muitos sistemas no mundo real compostos por partes individuais conectadas de alguma maneira. Uma rede é uma das formas mais simples e poderosas de representar o padrão de conexão ou a interação entre as partes desses sistemas. (9) Em uma rede, as partes ou componentes do sistema são tomados como vértices, enquanto as relações entre as partes são as arestas. Essa simples representação abstrata reduz sistemas complexos a um conjunto básico de variáveis (*e.g.*, grau, distância geodésica, centralidade) que capturam o padrão de conexão entre os componentes.

Alguns exemplos de sistemas que podem ser representados por redes são:

1. Internet: os vértices são os computadores e as arestas são as conexões físicas entre os computadores, seja conexões via cabo ou *wireless*.
2. *World wide web*: também chamada somente de “*web*” . Embora seja muitas vezes confundida com a internet, ambos são conceitos distintos. A internet representa a conexão física entre computadores, enquanto a web refere-se a páginas ou *sites*. Nesta rede os vértices são as páginas e as arestas os *hiperlinks*.
3. Redes sociais: são redes de pessoas e as conexões podem ter diversos significados em variados contextos. As redes sociais englobam as redes de amizades, redes de funcionários de uma companhia, redes de colaboração científica, grupos de pessoas com interesses ou práticas em comum, etc. Nesse tipo de rede, os vértices são as pessoas e as arestas representam o tipo de relação entre elas.
4. Redes ecológicas: as teias alimentares são um tipo de rede, onde os vértices são os animais e as arestas a relação de predação entre eles.
5. Redes metabólicas: representam um conjunto de reações químicas em seres vivos, onde os vértices são metabólitos e as arestas são reações bioquímicas.
6. Redes neurais: cada vértice é um neurônio e as arestas, as sinapses.
7. Redes de infraestrutura: nessa classe de redes estão, por exemplo, as redes de transmissão de eletricidade, rede de aeroportos e redes de rodovias.

Existem muitos outros exemplos além dos citados acima. As redes estão presentes em diversas áreas do conhecimento como física, matemática, biologia, ciências sociais, ciências da computação, linguística, engenharia, entre outras áreas.

A estrutura da rede, bem como o padrão de conexão entre os vértices, podem ter grande influência no comportamento da rede. Por exemplo, em redes sociais o aumento da conectividade e a proximidade espacial pode facilitar a disseminação de epidemias ou de informações falsas. (5) Em redes de internet, a estrutura e a forma como os vértices estão conectados pode interferir no tráfego de pacotes. Por essas razões, os cientistas têm desenvolvido através dos anos métodos matemático-computacionais e métodos estatísticos para modelar e estudar o comportamento de redes e, por consequência, classificar redes, encontrar padrões que diferenciam uma rede de outra. Por esses dois motivos; 1) Sistemas reais podem ser representados por redes e, 2) a estrutura da rede está fortemente ligada ao comportamento da mesma, que a área de reconhecimento de padrões tornou-se imprescindível na ciência de redes.

1.2 Objetivo

O objetivo deste trabalho é reconhecer padrões em modelos de redes e em redes do mundo real usando uma variação do método da caminhada determinística do turista. O reconhecimento de padrões em redes complexas apresentado neste estudo consiste de duas abordagens:

1. Desenvolvimento de uma métrica baseada no algoritmo das caminhadas determinísticas do turista capaz de reconhecer redes com propriedades de pequeno-mundo.
2. Proposição de um método para classificação de redes reais e sintéticas baseado nas caminhadas determinísticas do turista.

Por meio dessas abordagens, busca-se contribuir para a compreensão e caracterização de redes complexas.

2 CIÊNCIA DE REDES

A ciência de redes surgiu visando entender e modelar sistemas do mundo real que podem ser descritos na forma de uma rede, isto é, sistemas que possuem partes individuais conectadas através de alguma relação, a qual muitas vezes é não trivial. As redes estão em todos os lugares, da internet ao transporte, das redes sociais à transmissão de doenças. Muitos modelos emergiram durante as últimas décadas para modelar sistemas reais e tentar compreender sua natureza através de medidas de redes, tais como grau, coeficiente de aglomeração, distância, entre outras. (10) A Seção 2.2 mostra alguns desses modelos.

Uma rede é um grafo $G(V, E)$ composto por um conjunto V de vértices ou nós e um conjunto de arestas E . A rede pode ser não dirigida (ver Figura 2.(a)), onde se o nó i está conectado ao nó j , o oposto também é verdade, j está conectado a i , ou a rede pode ser dirigida (Figura 2.(b)), portanto, se i está conectado com j , a direção oposta da conexão pode não existir. Grafos dirigidos também são chamados de digrafos. De forma matemática, uma aresta é representada por uma tupla $(i, j) \in E$, indicando uma conexão partindo do vértice $i \in V$ ao vértice $j \in V$. Cada conexão pode ter um peso, como mostra a Figura 2.(c). Na seção a seguir veremos algumas propriedades das redes.

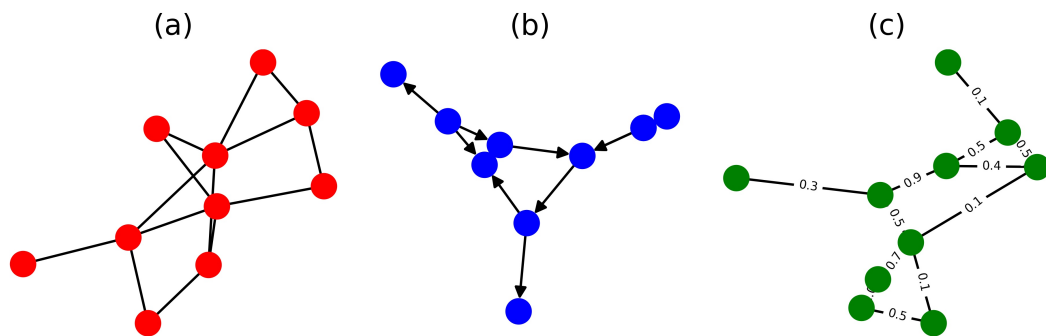


Figura 2 – Tipos de grafos. (a) Grafo não dirigido. (b) Grafo dirigido (digrafo). (c) Grafo com pesos.

Fonte: Elaborada pelo autor.

2.1 Propriedades e medidas de redes

Suponha uma rede $G(V, E)$ com n vértices e m arestas, onde cada aresta é representada por uma tupla (i, j) . Então seguem as seguintes propriedades:

1. O vértice $i \in V$ é dito adjacente ao vértice $j \in V$ se houver uma aresta conectando ambos, isto é, $(i, j) \in E$.

2. O conjunto vizinhança U_i associado ao nó i é o conjunto de todos os nós adjacentes a i . Para grafos dirigidos, U_i pode ser trocado pelo conjunto dos nós sucessores ou predecessores à i .
3. Uma lista de arestas $L_E = [(v_1, v_2), (v_1, v_3), \dots]$, onde cada entrada é uma tupla representando uma aresta, é uma forma simples de representar um grafo e de simples armazenamento em memória.
4. Uma lista de adjacências é outra forma de armazenar um grafo computacionalmente, a qual a entrada é o conjunto vizinhança U_i associado ao i -ésimo nó: $L_A = [U_1, U_2, U_3, \dots, U_i, \dots, U_n]$. Pode ser representada computacionalmente por um dicionário ou tabela hash.
5. A matriz de adjacências (A) é uma alternativa à lista de adjacências. Trata-se de uma matriz de dimensão $n \times n$ que contém todas as adjacências do grafo, onde o elemento $A_{ij} = 1$ se houver adjacência entre os vértices i e j e $A_{ij} = 0$ no caso contrário. De forma matemática:

$$A_{ij} = \begin{cases} 1 & , \text{ Se } (i, j) \in E \\ 0 & \text{ do contrário} \end{cases} \quad (2.1)$$

Os elementos da diagonal principal ($A_{ii} \forall i \in V$) são todos nulos, exceto se houver *self-loops*, onde o vértice fonte e o vértice alvo são o mesmo vértice.

Se a rede for não-dirigida, então $A_{ij} = A_{ji}$, o que faz de A uma matriz simétrica em relação à diagonal. Para grafos dirigidos, não há simetria na matriz de adjacência e ela pode ser substituída por uma matriz de incidência.

A desvantagem das matrizes de adjacência em relação às listas de adjacências é que no primeiro caso, mesmo que não haja conexão entre um dado vértice i e um dado vértice j , ainda assim o elemento A_{ij} assume um valor, no caso, $A_{ij} = 0$, e toma espaço de memória. Se o grafo for esparsos, a lista de adjacência é, portanto, mais eficiente em relação ao gasto de memória, representando somente as conexões que de fato estão no grafo.

6. Uma rede pesada é um grafo o qual as arestas possuem um peso. Desta forma, a matriz de adjacência é:

$$A_{ij} = \begin{cases} w_{ij} & , \text{ Se } (i, j) \in E \\ 0 & \text{ do contrário} \end{cases} \quad (2.2)$$

onde w_{ij} é o peso associado a aresta (i, j) .

7. O grau de um vértice i é a soma de todas as arestas que passam por ele. Podemos representar essa soma em termos da matriz de adjacência:

$$k_i = \sum_{j=1}^n A_{ij} \quad (2.3)$$

8. Para redes dirigidas, o grau é substituído pelo grau de saída e/ou pelo grau de entrada que medem a quantidade de arestas que saem ou entram no vértice, respectivamente. Os graus de entrada e de saída podem ser escritos como:

$$k_i^{in} = \sum_{j=1}^n A_{ji}, \quad k_i^{out} = \sum_{j=1}^n A_{ij} \quad (2.4)$$

9. O grau médio da rede é definido como:

$$\langle k \rangle = \frac{1}{n} \sum_{i=1}^n k_i \quad (2.5)$$

Para redes com m arestas, n vértices e conexões não dirigidas, o grau médio da rede pode ser calculado pela seguinte razão:

$$\langle k \rangle = \frac{2m}{n} \quad (2.6)$$

Isso decorre de que em um grafo não dirigido com m arestas cada aresta possui duas pontas, portanto, são $2m$ pontas de arestas em todo o grafo. Mas o número de pontas equivale ao grau total do grafo: $\sum_i k_i = 2m$ que, normalizando pelo tamanho da rede, resulta na Equação 2.6.

10. Um caminho é uma sequência de vértices onde cada par consecutivo de vértices é conectado por uma aresta. Um caminho que termina no vértice inicial é chamado de ciclo.
11. Um grafo é dito conexo se existir pelo menos um caminho entre cada par de nós do grafo. Caso contrário, o grafo é chamado de desconexo.
12. O comprimento de um caminho é a quantidade de arestas presentes ao longo dele. Uma mesma aresta pode ser computada mais de uma vez, quando o caminho atravessar pela mesma aresta.
13. Pode haver muitos caminhos conectando um mesmo par de vértices, no entanto, o menor desses caminhos é chamado de caminho geodésico.
14. A distância entre dois nós i e j , também chamada de distância geodésica e denotada por $d(i, j)$, é o comprimento do menor caminho entre i e j , isto é, o comprimento do

caminho geodésico entre i e j . Em 1956 Edsger Dijkstra desenvolveu um algoritmo para calcular a distância entre pares de nós.

15. O diâmetro da rede é o comprimento do mais longo caminho geodésico contido na rede.
16. A transitividade quantifica a ocorrência de triângulos em uma rede. Seja u, v, w três nós da rede G . A transitividade é 1 se u está ligado com v , u está ligado à w e v está ligado à w , formando, portanto, um ciclo triangular entre eles. Essa é uma propriedade bastante presente em redes sociais, onde se a pessoa u conhece a pessoa v e a pessoa v conhece a pessoa w , há grande chance de u conhecer w , ou em outras palavras, amigo de amigo é também um amigo. Uma maneira de medir a transitividade de uma rede é através do coeficiente de aglomeração ou *Clustering coefficient*, em inglês.
17. O coeficiente de aglomeração, ou *clustering coefficient*, em inglês, é definido como:

$$C = \frac{3 \times \text{número de triângulos}}{\text{Número de triplas conectadas}} \quad (2.7)$$

O coeficiente de aglomeração, também chamado de coeficiente de aglomeração global, mede a razão da quantidade de triângulos na rede com a quantidade de triplas de vértices conectadas. Onde uma tripla conectada significa que, dado três vértices u, v e w , u está conectado com v e v está conectado com w . A conexão entre u e w pode existir ou não. De forma equivalente, podemos dizer que há um caminho de comprimento 2 conectando u à w .

18. O coeficiente de aglomeração local associado ao nó i é:

$$C_i = \frac{\text{quantidades de pares de vizinhos do nó } i \text{ conectados}}{\text{quantidade de pares de vizinhos do nó } i} \quad (2.8)$$

19. O coeficiente de aglomeração local médio é:

$$\langle C_i \rangle = \frac{1}{n} \sum_{i=1}^n C_i \quad (2.9)$$

onde n é a quantidade de nós.

Embora $\langle C_i \rangle$ seja uma medida global de coeficiente de aglomeração, tal como C , existe diferença entre as duas medidas. A primeira trata todos os vértices com o mesmo peso, enquanto a segunda trata todos os triângulos com o mesmo peso, sendo assim, embora sejam duas formas de medir a transitividade de uma rede, há divergência no resultado delas. (9–10)

20. A assortatividade é uma métrica que quantifica a tendência de vértices a se conectarem àqueles que são de, alguma forma, semelhantes. Por exemplo, a assortatividade do grau mede a tendência de vértices a se ligarem a vértices com grau similar. Existem algumas formas de calcular a assortatividade, uma delas é usando o coeficiente de correlação de Pearson (r):

$$r = \frac{S_1 S_e - S_2^2}{S_1 S_3 - S_2^2} \quad (2.10)$$

Sendo,

$$S_1 = \sum_i k_i, \quad S_2 = \sum_i k_i^2, \quad S_3 = \sum_i k_i^3, \quad S_e = \sum_{ij} A_{ij} k_i k_j \quad (2.11)$$

Se $r > 0$, os nós tendem a se conectarem com nós de grau similar. Se $r < 0$, os nós tendem a se conectarem com nós de graus diferentes. Se $r \approx 0$, os nós não tem preferência.

2.1.1 Medidas de centralidade

A centralidade mede o quanto que um vértice é importante para a disseminação de informação na rede. Para ilustrar essa questão, tomemos como exemplo o aeroporto de Congonhas em São Paulo. Este aeroporto possui maior fluxo de passageiros e uma maior quantidade de linhas aéreas do que o aeroporto de uma cidade mediana. Portanto, uma falha no aeroporto de Congonhas causaria maiores transtornos, podendo, inclusive, afetar dezenas de outros aeroportos e centenas de outras linhas em efeito cascata. Então, medir a centralidade de vértices é fundamental para compreender a estrutura e o comportamento da rede. Existem diversas maneiras de se medir a centralidade e alguns desses métodos são apresentados abaixo.

1. A centralidade de grau (*degree centrality*) é uma medida de centralidade baseada no valor de grau dos vértices. Quanto maior o grau, maior a tendência do vértice de concentrar o fluxo de informação pela rede, bem como de propagar mais rápido a informação. A maior limitação dessa medida é a existência de *hubs* periféricos; nós com muitas conexões, porém localizados em partes periféricas da rede, não propagando rapidamente informação para fora de seu grupo local. Sendo assim, é uma boa medida para centralidade local, mas não global.
2. A *closeness centrality* é uma medida de centralidade que considera a distância geodésica de um dado nó i aos demais nós. Essa medida é definida como:

$$CL_i = \frac{n}{\sum_{j=1}^n d_{ij}} \quad (2.12)$$

De acordo com essa medida, vértices que possuem distância curta em relação a todos os outros vértices é considerado um vértice central. É uma medida fácil de computar, porém, sua maior limitação é o estreito intervalo de variação de rede para rede, uma vez que, a maioria das redes possui distâncias entre nós não muito elevadas.

3. A *Betweenness centrality* mede a quantidade de tráfego que um vértice recebe quando a informação sai de partes distintas da rede, em outras palavras, essa medida verifica a quantidade de vezes que o nó age como uma ponte, conectando diversos setores da rede. Essa medida é definida como:

$$B_i = \sum_{(a,b)} \frac{\eta(a,i,b)}{\eta(a,b)} \quad (2.13)$$

onde $\eta(a,b)$ é a quantidade de caminhos geodésicos entre os nós a e b e $\eta(a,i,b)$ é a quantidade de caminhos geodésicos entre a e b que passam pelo vértice i . Sua limitação é o alto custo computacional.

4. A *Eigenvector centrality* baseia-se no processo de difusão de informação pela rede. Seja o vetor de centralidade X contendo a centralidade x_i de todos os nós da rede. Tomemos inicialmente (X_0) , com todas as entradas desse vetor igual a um, *i.e.*, $x_i = 1 \forall x_i \in X_0$. Esse vetor será multiplicado pela matriz de adjacência a cada iteração e após t iterações esse vetor será:

$$X_t = X_{t-1} \cdot A \quad (2.14)$$

onde X_t é o vetor após t iterações, X_{t-1} é o vetor após $t - 1$ iterações e A é a matriz de adjacência. Sabendo que esse processo consiste da multiplicação consecutiva de X pela matriz de adjacência, podemos simplificar tomando $X_t = X_0 A^t$. Podemos escrever X_0 como combinação linear dos autovetores v_k de A :

$$X_0 = \sum_k c_k v_k \quad (2.15)$$

com c_k escolhidos de forma apropriada. Temos, portanto:

$$X_t = A^t \sum_k c_k v_k \quad (2.16)$$

Sabendo que $Av_k = \lambda_k v_k$, então:

$$X_t = \sum_k c_k \lambda_k^t v_k \quad (2.17)$$

Em geral, haverá muitos autovalores diferentes que geram autovetores X_t não nulos. No entanto, desejamos que todas as entradas de X_t sejam positivas, implicando, pelo

teorema de Perron-Frobenius, que somente o maior autovalor (λ_1) seja considerado. Multiplicando e dividindo a Equação 2.17 por λ_1 , temos:

$$X_t = \lambda_1^t \sum_k c_k \frac{\lambda_k^t}{\lambda_1^t} v_k \quad (2.18)$$

Como essa soma decai exponencialmente conforme t aumenta, consideramos somente o maior termo da soma $X_t = c_1 \lambda_1^t v_1$ associado ao maior autovalor de A .

As entradas de X_t serão os *scores* de centralidade de cada nó da rede. Os nós com maiores *scores* são aqueles que acumularam mais informações, portanto, são os nós centrais da rede.

5. Uma outra medida de centralidade é a *Acessibilidade*. (11) Seja uma rede com n nós e $P_t(i, j)$ a probabilidade de transição que um caminhante partindo do nó j tem de alcançar o nó i após t passos. Para estimar tal probabilidade, M caminhadas com o mesmo ponto inicial j são realizadas e então a quantidade de vezes em que o caminhante atinge o nó i é medida e dividida pelo total M de caminhadas. Tendo $P_t(i, j)$, para todos os nós iniciais j , é calculado a entropia de Shannon (H) quantificando a diversidade de caminhadas que passam por i :

$$H_t(i, \Omega) = - \sum_{j=1, j \neq i}^n P_t(i, j) \log(P_t(i, j)) \quad (2.19)$$

onde Ω é o conjunto de todos os nós da rede exceto o nó i . Então a acessibilidade associada ao nó i é definida em termos da entropia de Shannon H_t e normalizada pelos $n - 1$ nós restantes:

$$A_t(i) = \frac{e^{H_t}}{n - 1} \quad (2.20)$$

A representação gráfica das medidas de centralidade definidas acima é mostrada na Figura 3.

2.2 Modelos de redes

O principal objetivo dos modelos teóricos de redes é prever a evolução de redes do mundo real, *e.g.*, a internet, redes biológicas e redes sociais. Os modelos teóricos permitem gerar redes com propriedades que podem ser controladas, tais como o grau, o coeficiente de aglomeração, etc. Então, o uso desses modelos facilita o estudo de como a estrutura da rede influencia na dinâmica, sendo bastante útil para compreender processos como disseminação de doenças e notícias falsas. Nessa seção serão apresentados alguns dos mais importantes modelos de redes.

2.2.1 Redes regulares

As redes regulares são o modelo mais básico de redes. Elas, em geral, possuem um alto coeficiente de aglomeração e um elevado diâmetro. Sua principal característica é que todos os vértices possuem o mesmo grau, portanto, a distribuição de grau é um delta de Dirac.

Um exemplo de rede regular é ilustrado pela Figura 4(a). Uma rede unidimensional circular com condição periódica de contorno e grau médio $\langle k \rangle = k$. Para essa rede o coeficiente de aglomeração vale:

$$C = \frac{3(\langle k \rangle - 2)}{4(\langle k \rangle - 1)} \quad (2.21)$$

onde no limite de grau alto, o coeficiente tende à $\frac{3}{4}$, um valor considerado alto. A média dos menores caminhos vale: $\langle \ell \rangle = \frac{n}{2\langle k \rangle}$, aumentando linearmente em relação ao tamanho da rede, implicando que conforme a rede torna-se maior, a distância entre quaisquer nós da rede atinge valores elevados. A demonstração desses resultados está no livro “*Networks, an introduction*” de M. E. J. Newman, quarta edição, páginas 552 à 555. (9)

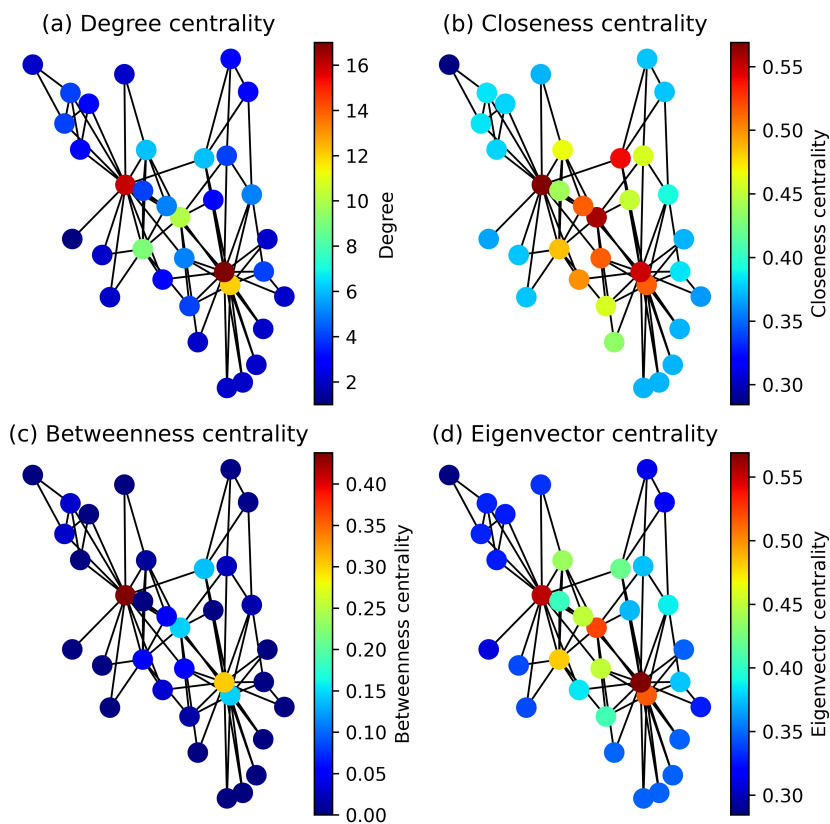


Figura 3 – Medidas de centralidade.

Fonte: Elaborada pelo autor.

2.2.2 Redes aleatórias

O modelo de redes aleatórias é um dos mais simples e um dos mais antigos modelos de rede. As redes aleatórias foram introduzidas em 1951 por Ray J. Solomonoff e Anatol Rapoport (5) e reintroduzidas por E. N. Gilbert. (7) Foram estudadas mais profundamente por Paul Erdos e Alfred Renyi, os quais popularizaram o modelo. (6)

Erdos e Renyi definiram um algoritmo para gerar um *ensemble* de redes aleatórias, denotado por $G(n, m)$, onde tomando uma rede com n vértices, m arestas são distribuídas de forma aleatória sobre pares distintos de vértices. Neste modelo, chamado de modelo de Erdos-Renyi (ER), a quantidade de vértices e a quantidade de arestas são parâmetros controlados e é um modelo simples de construir. Contemporaneamente e, de forma independente, Edgar Gilbert criou outro algoritmo denotado por $G(n, p)$ e referenciado algumas vezes por modelo Erdos-Renyi-Gilbert. Neste modelo, as arestas são adicionadas aleatoriamente entre pares distintos de nós seguindo uma probabilidade de conexão p , formando, portanto, um processo de Bernoulli com probabilidade de conexão p e probabilidade de não-conexão igual a $1 - p$. Se p é zero, então não há nenhuma conexão na rede, enquanto se p for igual a 1, então todos os nós estão conectados com os outros $n - 1$ nós. Nesse último caso, todos os vértices assumem o mesmo valor de grau, $k = n - 1$ e, portanto, a rede perde a aleatoriedade e torna-se regular (ver Seção 2.2.1).

Ambos os algoritmos geram *ensemble* de redes aleatórias, uma vez que, cada execução desses algoritmos gera redes diferentes, cada uma com um conjunto de arestas diferente. No grafo $G(n, m)$ a quantidade de arestas é um parâmetro fixo, no grafo $G(n, p)$ isso não ocorre. Considerando a partir de agora o último modelo e dado que cada aresta conecta um par de vértices, então a probabilidade de gerar grafos com exatos n vértices e m arestas formados a partir de $\binom{n}{2}$ pares distintos de vértices é:

$$P(m) = \binom{\binom{n}{2}}{m} p^m (1 - p)^{\binom{n}{2} - m} \quad (2.22)$$

o qual é a distribuição binomial, portanto, a quantidade média de arestas vale:

$$\langle m \rangle = \sum_{m=0}^{\binom{n}{2}} m P(m) = \binom{n}{2} p \quad (2.23)$$

Dada a Eq. 2.6, o grau médio de uma rede $G(n, p)$ é:

$$\langle k \rangle = \sum_{m=0}^{\binom{n}{2}} \frac{2m}{n} P(m) = (n - 1)p \quad (2.24)$$

e a distribuição de grau é:

$$P_k = \binom{n-1}{k} p^k (1-p)^{n-1-k} \quad (2.25)$$

a qual é uma distribuição de binomial. Tomando o limite para n grande temos:

$$P_k = \frac{\langle k \rangle^k}{k!} e^{-\langle k \rangle} \quad (2.26)$$

uma distribuição de Poisson (9), como mostra a Figura 7.(a). O coeficiente de aglomeração de uma rede aleatória é:

$$C = p = \frac{\langle k \rangle}{n-1} \quad (2.27)$$

implicando que para redes grandes, o coeficiente de aglomeração é baixo. O diâmetro de um grafo aleatório vale:

$$D = \frac{\ln n}{\ln \langle k \rangle} \quad (2.28)$$

Note que o diâmetro cresce suavemente com o aumento do tamanho da rede. Mesmo para redes grandes, com milhões ou até bilhões de vértices, o diâmetro da rede permanece pequeno. Esse efeito é chamado de efeito de pequeno-mundo e que será discutido mais profundamente na Seção 2.2.3.

Embora simples e amplamente estudado, o modelo de redes aleatórias oferece uma aproximação pobre das redes do mundo real, divergindo das redes reais tanto na distribuição de grau, quanto no coeficiente de aglomeração. Os modelos a seguir são mais acurados para descrever redes do mundo real.

2.2.3 Redes de pequeno-mundo

As duas seções anteriores mostraram que redes regulares possuem alto coeficiente de aglomeração e alta distância entre vértices, enquanto redes aleatórias possuem baixo coeficiente de aglomeração e baixa distância entre vértices. No entanto, muitas redes do mundo real, especialmente as redes sociais, possuem baixa distância entre vértices, o efeito pequeno-mundo, e possuem alto coeficiente de aglomeração que, em linguagem de redes sociais, significa que amigo de amigo é também um amigo. (12, 13)

O efeito de pequeno-mundo tem sido exaustivamente estudado a partir da segunda metade do século XX, principalmente na área das ciências sociais, uma vez que, redes sociais frequentemente apresentam esse fenômeno. Por motivos históricos vale citar o experimento de pequeno-mundo executado por Stanley Milgram entre as décadas de 1960 e 1970. (14) O experimento consistia em enviar cartas para um conjunto de pessoas aleatórias em cidades distantes dos Estados Unidos da América e cada carta continha um nome. Se o destinatário conhecesse pessoalmente a pessoa citada, então reenviava a carta diretamente

para a tal pessoa. Caso contrário, ela reenviava a carta para alguém a qual ela supunha que talvez conhecesse a pessoa. O objetivo desse experimento era quantificar o grau de separação entre quaisquer duas pessoas no mundo, o que levou à hipótese dos “seis graus de separação”, onde cada pessoa estaria afastada de qualquer outra, em média, por uma distância igual a seis. (15)

Em razão de tentar modelar redes que apresentam o efeito de pequeno-mundo e são altamente “clusterizadas”, Duncan J. Watts e Steven H. Strogatz construíram um algoritmo capaz de produzir redes com tais características, as redes de pequeno-mundo. (16)

O modelo WS começa com uma rede regular circular com condição de contorno periódica, conforme mostra a Figura 4.(a). Em seguida, cada aresta da rede tem uma das pontas religada a outro nó da rede de acordo com uma probabilidade de religação p . Dessa forma, se $p = 0$ nenhuma religação é feita e a rede permanece regular, com alto coeficiente de aglomeração e alta distância entre nós (ver Figura 4.(a)). Para $p = 1$ (ver Figura 4.(c)) a rede torna-se plenamente aleatória, com distribuição de grau semelhante a de uma rede ER e tanto o coeficiente de aglomeração, quanto o diâmetro assumem valores baixos. Para uma certa faixa de probabilidade $0 < p < 1$ (ver Figura 4.(b)), a rede gerada possui, simultaneamente, alto coeficiente de aglomeração e baixo diâmetro semelhante a uma rede real. O processo de religação de arestas cria “atalhos” na rede, conectando diretamente vértices que estavam anteriormente conectados através de um longo caminho. O comprimento dos menores caminhos, *i.e.*, a distância geodésica, decresce rapidamente com p , enquanto o coeficiente de aglomeração mantém-se alto, mesmo para altos valores de p , como mostra a Figura 4(d), então, para um p intermediário, a rede será uma rede de pequeno-mundo.

Nenhuma aresta foi adicionada ou removida da rede, somente religada, portanto, o grau médio é o mesmo da rede regular inicial, porém a distribuição de grau muda, tendendo à distribuição de Poisson quando $p \rightarrow 1$:

$$P_k = \frac{(\langle k \rangle p)^{k - \langle k \rangle}}{(k - \langle k \rangle)!} e^{-\langle k \rangle p} \quad (2.29)$$

tornando-se, nessa propriedade, semelhante ao modelo ER.

Existe um outro modelo de redes de pequeno-mundo. Em uma variação do modelo WS, apresentada por M. E. J. Newman (9), ao invés de arestas serem religadas, novas arestas são adicionadas à rede de acordo com uma probabilidade p . Dessa forma o grau médio da rede varia, porém, o coeficiente de aglomeração e o comprimento dos menores caminhos seguem o mesmo padrão do modelo WS original.

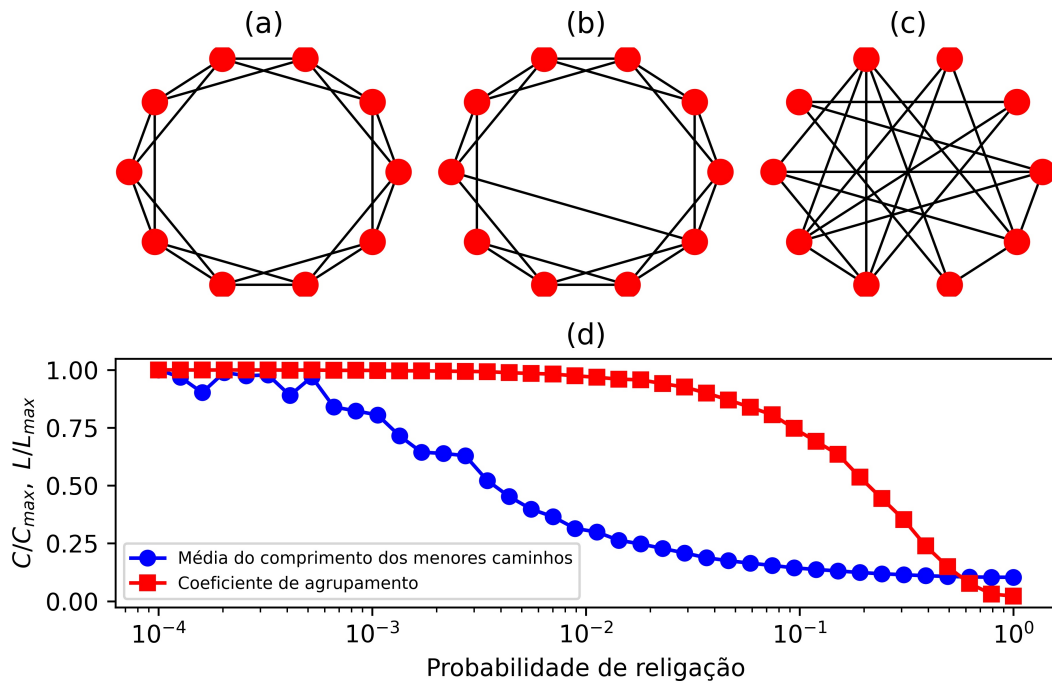


Figura 4 – Modelo de Watts-Strogatz. (a) Rede regular. (b) Rede de pequeno-mundo. (c) Rede aleatória. (d) Curvas de média do comprimento dos menores caminhos (azul) e coeficiente de aglomeração global (vermelho).

Fonte: Elaborada pelo autor.

O coeficiente de aglomeração para a rede de pequeno-mundo vale:

$$C = \frac{3(\langle k \rangle - 2)}{4(\langle k \rangle - 1) + 8\langle k \rangle p + 4\langle k \rangle p^2} \quad (2.30)$$

retornando para a Equação 2.21 se $p = 0$.

O diâmetro da rede é:

$$D = \frac{\ln(n\langle k \rangle p)}{\langle k \rangle^2 p} \quad (2.31)$$

o qual cresce suavemente com o tamanho da rede, semelhante ao diâmetro de uma rede aleatória dada pela Equação 2.28.

Uma vez que muitas redes do mundo real apresentam as propriedades de rede de pequeno-mundo, encontrar uma métrica que quantifique o quão próximo uma rede está de uma rede de pequeno-mundo tornou-se crucial. Em 1999, Walsh *et. al.* apresentaram o “raio de proximidade” (17), definido como:

$$\sigma = \frac{\frac{C}{C_{random}}}{\frac{L}{L_{random}}} \quad (2.32)$$

Onde L e C são o comprimento médio dos menores caminhos e o coeficiente de aglomeração, respectivamente. Os termos L_{random} e C_{random} referem-se as mesmas medidas, mas para redes aleatórias equivalentes. Se $L/L_{random} \approx 1$ e $C/C_{random} \gg 1$, então a rede se assemelha a uma rede de pequeno-mundo.

Uma abordagem alternativa foi introduzida por Telesford *et. al.* (18) e definido como:

$$\omega = \frac{L_{random}}{L} - \frac{C}{C_{lattice}} \quad (2.33)$$

Onde L_{random} e $C_{lattice}$ são o diâmetro de uma rede aleatória equivalente e o coeficiente de aglomeração para uma rede regular equivalente, respectivamente. Se $\omega \rightarrow 1$ a rede tem comportamento de rede aleatória, se $\omega \rightarrow -1$ a rede tem alta regularidade e para $\omega \rightarrow 0$ a rede se assemelha a uma rede de pequeno-mundo. Embora o valor $\omega = 0$ seja o valor ideal que uma rede deveria apresentar para ser chamada de pequeno-mundo, há um intervalo em torno de $\omega = 0$, considerado pelos autores do artigo, em que a rede apresentaria comportamentos de rede de pequeno-mundo. Os autores indicaram que no intervalo $-0.5 \leq \omega \leq 0.5$, a rede exibe comportamentos de pequeno-mundo.

2.2.4 Redes livre de escala

Os modelos de pequeno-mundo emergiram para suprir uma falha do modelo de redes aleatórias no que se refere a alta transitividade e baixo diâmetro das redes do mundo real. Entretanto, ambos os modelos falham em simular a distribuição de grau de redes reais. Tanto as redes aleatórias quanto as redes de pequeno-mundo possuem distribuição de grau seguindo a distribuição de Poisson (ver Fig. 7.(a) e Fig. 7.(b)), porém, inúmeros sistemas reais apresentam uma distribuição de grau que segue uma lei de potência (ou distribuição de Pareto) (19–21), como mostra a figura 5 e cuja fórmula é:

$$P(k) \sim k^{-\gamma} \quad (2.34)$$

Os modelos que surgiram para imitar uma distribuição de grau de Pareto foram chamados de redes livres de escala (*scale-free networks*, em inglês) e esse modelo foi popularizado por László Barabási e Réka Albert (20), que desenvolveram o mais famoso modelo desse tipo, o modelo Barabási-Albert (BA), o qual será o foco dessa seção. No entanto, vale mencionar outros modelos como o modelo de Price (19), o de Dorogovtsev & Mendes (22) e o de Krapivsky *et. al.* (23)

O modelo de Barabási-Albert baseia-se em dois passos. No primeiro, chamado de crescimento, uma rede não dirigida com n nós, é criada e então vértices são adicionados com uma probabilidade uniforme p semelhantemente ao modelo ER. A segunda etapa consiste na ligação preferencial (*Preferential attachment*), onde os vértices ganham novas

conexões conforme a quantidade de conexões que eles já possuem. Quem tem mais conexões tende a ganhar mais e quem tem menos, tende a ganhar poucas conexões, o que, em jargão popular pode ser traduzido como “rico fica cada vez mais rico”. Os vértices com mais conexões na rede são chamados de *hubs*. De forma matemática, a probabilidade de um vértice i ganhar novas conexões é:

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j} \quad (2.35)$$

Note que a probabilidade é linear em relação ao grau. Existem modelos que tomam a probabilidade de ligação preferencial como não-linear, mas tais modelos não serão tratados aqui.

O grau médio k_i do nó i evolui no tempo de acordo com quantas conexões (m) são adicionadas à rede a cada passo de tempo t , sendo então:

$$\frac{dk_i}{dt} = m\Pi(k_i) = \frac{mk_i}{\sum_j k_j} \quad (2.36)$$

Considerando que a cada passo de tempo a rede ganha m novas conexões e que a rede é não-dirigida, então o grau total da rede em um tempo t pode ser escrito como $\sum_j k_j = 2mt$. Substituindo esse valor na Eq. 2.36, temos:

$$\frac{dk_i}{dt} = \frac{k_i}{2t} \quad (2.37)$$

Escrevendo, portanto, o grau médio do vértice i em função do tempo, temos:

$$k_i(t) = m \left(\frac{t}{t_i} \right)^{1/2} \quad (2.38)$$

onde t_i é o tempo em que o vértice i foi adicionado na rede. Note que o grau de i não depende do grau inicial dele, mas depende do tempo de criação dele t_i , o que implica que os nós mais velhos da rede tendem a acumular mais ligações e a se tornarem *hubs*. Tomando $k_i = m$ e $t = t_i$, então a distribuição de grau no tempo t é:

$$P(k) = - \left(\frac{dk_i}{dt} \right)^{-1} = \frac{2m^2t}{k^3} = 2m^2tk^{-3} \quad (2.39)$$

Então a distribuição de grau depende da terceira potência do grau e, comparando com a Eq. 2.34 da lei de potência, temos que para o modelo de Barabási-Albert, $\gamma = 3$. A Figura 7.(d) mostra a distribuição de grau do modelo BA em escala logarítmica e note a semelhança com as distribuições de grau de algumas redes reais, ilustrado na Figura 5.

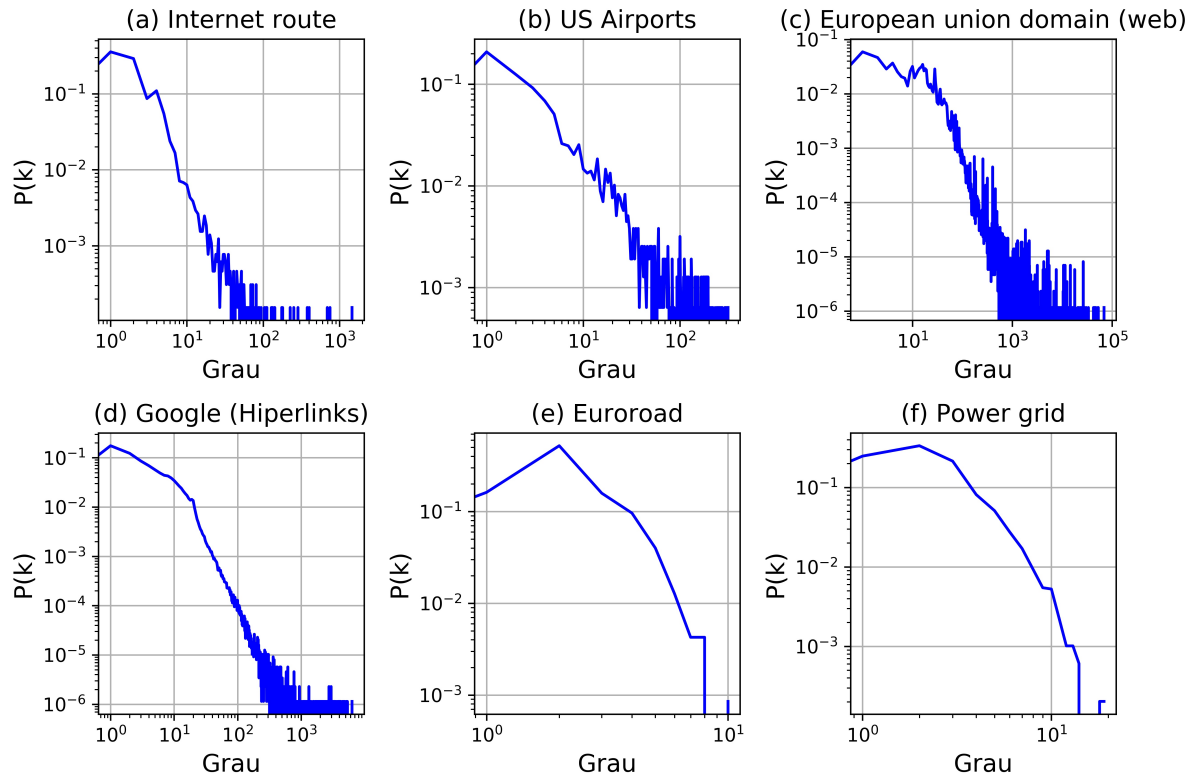


Figura 5 – Distribuição de grau de redes do mundo real.

Fonte: KONECT PROJECT (24)

2.2.5 Redes geométricas

Redes geométricas, também chamadas de redes espaciais ou ainda de redes geográficas, são redes as quais cada nó possui uma posição espacial bem definida. Muitas redes do mundo real são redes espaciais, como a rede de cidades, linhas de transmissão de eletricidade, estradas, aeroportos, dentre outras. Redes geométricas são bastante ricas, pois além das informações topológicas como as vistas na Seção 2.1, adiciona informações geométricas, tais como posição espacial, distância euclidiana, entre outras. (25) Assim como os modelos anteriores, há diversos algoritmos para construir redes geométricas, sendo que três deles merecem destaque: os **Grafos geométricos aleatórios**, as redes de **Voronoi** e as redes de **Waxman**, todos eles estão ilustrados na Figura 6.

1. Os Grafos geométricos aleatórios ou *Random geometric graph* (RGG) (26) são definidos como: dado um nó i e um raio de influência R_i centrado no nó i , todos os nós j que estão dentro desse raio de influência possuem probabilidade p de se conectarem ao nó i . A Figura 6.(b) mostra um exemplo de rede geométrica aleatória.

2. Redes de Voronoi: Dado um conjunto de vértices em uma região chamada de *Tesselação de Voronoi*, onde cada vértice define uma *célula de Voronoi*, então dois vértices são conectados se suas células de Voronoi forem adjacentes (ver Fig. 6.(a)). Esse processo, também chamado de *triangulação de Delaunay*, gera redes com topologias bem distintas dos grafos anteriores. (27)
3. As redes de Waxman (28) são semelhantes aos grafos geométricos aleatórios, mas a probabilidade de conexão de i com j decai exponencialmente com a distância euclidiana entre eles r_{ij} :

$$p_{ij} = e^{-\alpha r_{ij}} \quad (2.40)$$

onde α é um parâmetro de escala.

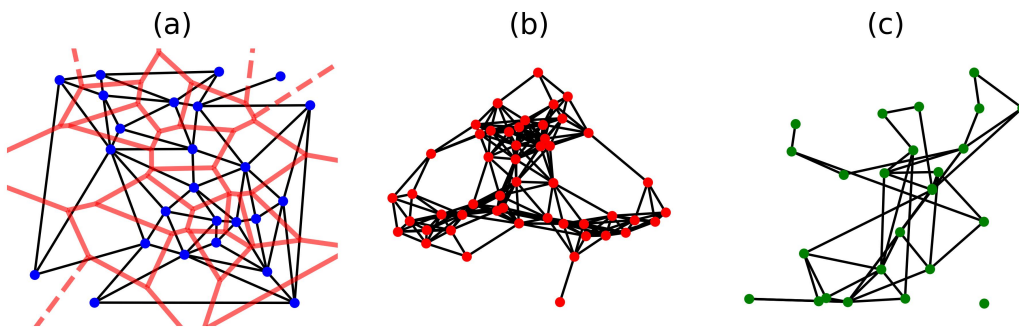


Figura 6 – Três modelos de redes geométricas. (a) Rede de Voronoi, incluindo as células de Voronoi. (b) Rede geométrica aleatória ($R = 0.25$). (c) Rede de Waxman ($\alpha = 0.1$).

Fonte: Elaborada pelo autor.

2.3 Métodos de reconhecimento de padrão em redes

Como visto anteriormente, reconhecer padrões em uma rede significa extrair um conjunto de atributos ou vetor de atributos que, dados como entrada em um algoritmo de aprendizado de máquinas, permita a classificação dessa rede. Diversos métodos surgiram nas últimas décadas e nessa seção será dada uma breve descrição de alguns deles, os quais serão relevantes para este trabalho.

2.3.1 Métodos baseados nas medidas e estruturas da rede

Os métodos mais tradicionais são, sem dúvida, os baseados em medidas topológicas da rede. Dado um conjunto de medidas Ω , é construído um vetor de atributos. Como há várias medidas, como mostrado na Seção 2.1, então há vários conjuntos de medidas

possíveis para serem tomados como vetor de atributos. A depender do *dataset*, um conjunto de medidas é mais vantajoso que outro, tornando o conjunto de medidas a ser usado um critério subjetivo. Há também a possibilidade de tomar medidas estatísticas a partir dessas medidas estruturais, como, por exemplo, o primeiro e o segundo momento estatístico, desvio padrão, entropia, etc.

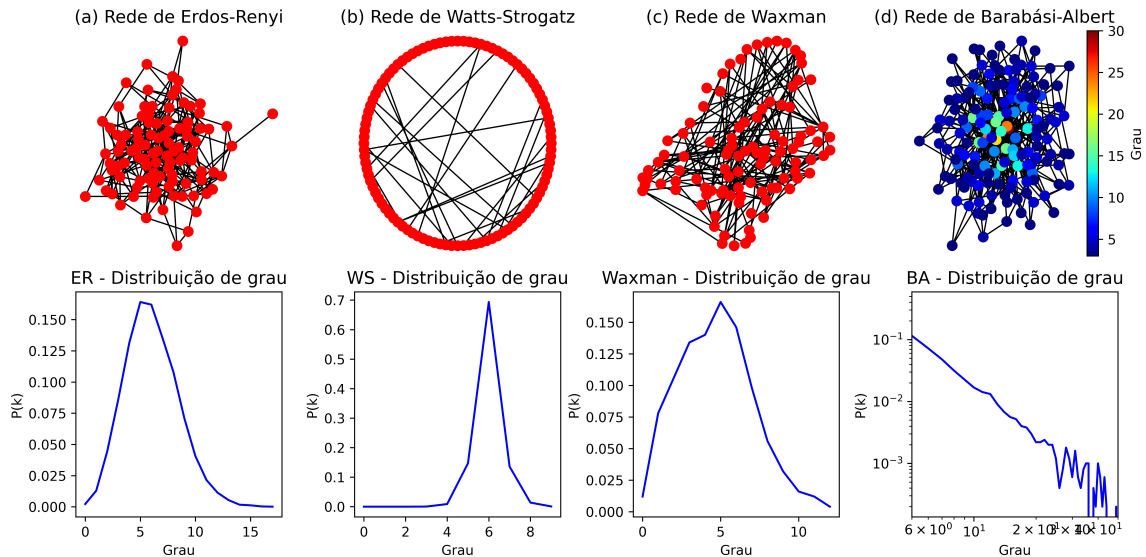


Figura 7 – Quatro modelos de redes e suas respectivas distribuições de grau.

Fonte: Elaborada pelo autor.

2.3.2 Métodos baseados em autômatos

Entre os métodos baseados em autômatos, três deles serão destacados abaixo.

1. ***Life-Like Network Automata (LLNA)***: O LLNA é um automata celular inspirado no *jogo da vida* de Conway (29), onde a vida e a morte de cada célula (representado tipicamente por 0 e 1) depende do ambiente, *i.e.*, de sua vizinhança. Em um artigo publicado em 2016, *Miranda et. al.* (30) propuseram o uso do padrão de evolução temporal (*Time evolution patterns* - TEP) como assinatura para sistemas complexos.

Dada uma tesselação τ contendo n células, onde cada célula é um vértice da rede conectado a k_i vizinhos e, dado um conjunto de estados iniciais S_0 , um conjunto de estados S e uma regra Φ , que governará a dinâmica das células, então o autômato R é definido como:

$$R = (\tau, S, S_0, \Phi) \quad (2.41)$$

Seja $s(c_i, t)$ o estado da célula c_i no tempo t , onde os valores de s estão restritos à $s = 0$ para morto e $s = 1$ para vivo. A transição de estado no tempo $t + 1$ é definido por $\Phi : s(c_i, t) \rightarrow s(c_i, t + 1)$, onde Φ é a regra de transição. Três transições são possíveis: se está morto, então nasce (*Born*) $0 \rightarrow 1$, se está vivo permanece vivo (*survive*) $1 \rightarrow 1$ e caso essa última situação não ocorrer, a célula morre. Cada transição depende da densidade $\sigma(c_i, t)$ de vizinhos vivos, onde:

$$\sigma(c_i, t) = \frac{1}{k_i} \sum_{j=1}^n A_{ij} S(c_j, t) \quad (2.42)$$

sendo A a matriz de adjacência da rede. Para ilustrar uma transição, tomemos como exemplo a regra B25-S8, onde B e S significam, respectivamente, *Born* e *Survive*. Seja uma célula c_i conectada com $r = 9$ vizinhos, incluindo a si mesma. Se a célula estiver morta ($s = 0$) e σ estiver entre X_x/r e $(X_x + 1)/r$, onde X_x é o x -ésimo dígito que segue a letra B (no exemplo, $X_1 = 2$ e $X_2 = 5$), então a célula realiza a transição $0 \rightarrow 1$. Se a célula estiver viva ($s = 1$) e σ estiver entre Y_y/r e $(Y_y + 1)/r$, onde Y_y é o y -ésimo dígito que segue a letra S (no exemplo $Y_1 = 8$), então a célula permanece viva. Usando nosso exemplo, temos:

$$0 \rightarrow 1 \begin{cases} \text{Se } \frac{2}{9} \leq \sigma(c_i, t) < \frac{3}{9} \\ \text{Se } \frac{5}{9} \leq \sigma(c_i, t) < \frac{6}{9} \end{cases} \quad (2.43)$$

$$1 \rightarrow 1 \begin{cases} \text{Se } \frac{8}{9} \leq \sigma(c_i, t) < \frac{9}{9} \end{cases} \quad (2.44)$$

A Equação 2.43 mostra a transição do tipo *Born* e a Equação 2.44 a transição do tipo *Survive*. Após t iterações, haverá uma sequência de estados para cada célula, *e.g.*, 011101101100101001 e, tomando a sequência de estados para cada célula, medidas podem ser realizadas como, por exemplo, a entropia ou a extração do padrão binário (LLNA-BP). (31) Esse método obteve alto desempenho no reconhecimento de padrões em modelos de redes e em redes do mundo real.

2. **Caminhadas aleatórias:** As caminhadas aleatórias, ou *Random walk*, são um tipo de autônomo onde o agente (ou caminhante) caminha sobre um grafo indo de vértice a vértice seguindo probabilidades de transição. Trata-se de um processo estocástico, um caso especial de cadeia de Markov, possuindo reversibilidade temporal, onde o caminho “de volta”, é também considerado uma caminhada aleatória. (32) Uma caminhada aleatória pode repetir vértices ou não e neste último caso ela se torna uma *Self-avoiding walk*. As caminhadas aleatórias foram inspiradas no movimento Browniano e podem ser usadas em diversos tipos de sistemas, como, por exemplo, para estudar flutuações de ações no mercado financeiro.

Seja $G(V, E)$ uma rede e considere um caminhante lançado a partir de um nó v_0 . A probabilidade de após t passos ele atingir o nó i é denotado por: $P_t(i) = Prob(v_t = i)$. Seja $M = (P_{ij})_{i,j \in V}$ a matriz estocástica de transição, onde cada probabilidade de transição P_{ij} é dada por:

$$p_{ij} = \begin{cases} \frac{1}{k_i}, & \text{Se } (i, j) \in E \\ 0, & \text{caso contrário} \end{cases} \quad (2.45)$$

A matriz de transição M também pode ser escrita em termos da matriz de adjacência: $M = DA$, onde A é matriz de adjacência da rede G e D a matriz diagonal, onde $D_{ii} = 1/k_i$. Então, a probabilidade de transição no passo $t + 1$ é:

$$P_{t+1} = M^T P_t \quad (2.46)$$

Podemos reescrever a Eq. 2.46 em termos da probabilidade inicial P_0 :

$$P_{t+1} = (M^T)^t P_0 \quad (2.47)$$

Lançando um caminhante de cada vértice da rede, teremos um conjunto de n trajetórias e medidas podem ser realizadas para caracterizar a rede, como a já citada acessibilidade.

3. **Caminhada determinística do turista:** A caminhada determinística do turista, ou *Deterministic Tourist walk* (DTW), é um autômato, desenvolvido originalmente para análise de textura em imagens (8), mas que pode ser inserido em diversos sistemas, incluindo redes. Assim como as caminhadas aleatórias, o algoritmo DTW é também um algoritmo de caminhada, entretanto há uma diferença fundamental entre os dois métodos; a caminhada do turista segue regras determinísticas e não está sujeito a uma matriz estocástica de transição. Podemos fazer a seguinte analogia: suponha um turista que quer visitar um conjunto de cidades em um país. Sua regra é deslocar-se para a cidade que apresenta menor custo de viagem (seja esse custo em termos de distância ou de dinheiro) a partir de sua cidade atual. Então a cada nova cidade o turista realiza o cálculo de qual a cidade vizinha que apresenta menor custo e a seleciona como destino. O método da caminhada do turista será profundamente abordado no Capítulo 3.

3 A CAMINHADA DETERMINÍSTICA DO TURISTA

A caminhada determinística do turista (DTW) é um método baseado em agentes desenvolvido no início do século XXI visando extrair assinaturas de textura em imagens. (8,33) Embora tenha sido inicialmente desenvolvido para a análise de textura, trata-se de um método bastante robusto e versátil, podendo ser aplicado também em redes. O método DTW trata-se de um autômato que se move por um meio, assim como a caminhada aleatória, porém se diferencia desta por seguir regras determinísticas de caminhada.

O método recebe este nome por se assemelhar a um turista que deseja visitar n cidade em um mapa, querendo sempre otimizar, de alguma forma, seu trajeto, seja seguindo sempre pela menor distância entre duas cidades, ou optando pelo menor custo financeiro, ou ainda seguindo uma “atratividade” pré-definida.

Há a possibilidade de o turista (ou caminhante) retornar para a cidade a qual visitou anteriormente. Sendo assim, é definido um tamanho de memória (μ) para o turista, de forma que ele não pode retornar a uma cidade armazenada em sua memória. Se o tamanho de memória for zero, então não há nenhuma restrição de cidade. Se o tamanho da memória for igual a um, então o turista se lembra da última cidade a qual visitou e evita retornar a ela. Caso o tamanho de memória seja grande o suficiente, o turista realizará a trajetória sem repetir nenhuma cidade, formando uma *self-avoiding walk*. Para tamanhos pequenos de memória, a caminhada do turista é parcialmente uma *self-avoiding walk*.

O turista realiza a caminhada sobre o mapa e sua trajetória é composta por duas partes, como mostra a Figura 8. A primeira parte é chamada de transiente e possui comprimento t , a segunda é o atrator, onde o turista entra numa região de tamanha uniformidade (ou atratividade) que fica preso em um ciclo de comprimento a encerrando, assim, sua trajetória. Devido à restrição imposta pela memória, o menor atrator possível possui comprimento $a = \mu + 1$. O comprimento total da trajetória do turista é a soma das duas partes: $\ell = t + a$.

Se a partir de cada uma das cidades do mapa for solto um turista, então ao fim haverá n trajetórias com diferentes combinações de transientes e atratores, levando a um histograma $h(\ell)$ da distribuição conjunta de transiente e atrator. Dado um comprimento ℓ de trajetória, o histograma é calculado da seguinte forma:

$$h(\ell) = \sum_{b=0}^{\ell} S(b, \ell - b) \quad (3.1)$$

onde $S(t, a)$ é a função contadora a qual o primeiro argumento é o comprimento do transiente e o segundo é o comprimento do atrator. A função S é definida como:

$$S(t, a) = \frac{1}{n} \sum_{i=1}^n \begin{cases} 1, & \text{Se } x_i = t \text{ e } y_i = a \\ 0, & \text{do contrário} \end{cases} \quad (3.2)$$

sendo n a quantidade de trajetórias.

Através do histograma podemos obter um vetor de características ϕ_μ^r para uma dada regra r e para um tamanho de memória μ . Tradicionalmente começamos esse vetor com o histograma para $\ell = \mu + 1$, o qual se refere ao menor comprimento possível do atrator. Então, o vetor ϕ é definido como:

$$\phi_\mu^r = [h(\ell = \mu + 1), h(\ell = \mu + 2), h(\ell = \mu + 3), \dots, h(\ell = \mu + z)] \quad (3.3)$$

onde $\mu + z$ é o *threshold* do tamanho da trajetória, sendo z um valor definido arbitrariamente.

Backes et. al. (34) mostraram ser possível aumentar o desempenho do método DTW concatenando o vetor da Eq. 3.3 para vários valores de memória, definindo assim um novo vetor $\psi_{\vec{\mu}}^r$:

$$\psi_{\vec{\mu}}^r = [\phi_{\mu_1}^r, \phi_{\mu_2}^r, \phi_{\mu_3}^r, \dots, \phi_{\mu_k}^r] \quad (3.4)$$

onde $\vec{\mu} = [\mu_1, \mu_2, \mu_3, \dots, \mu_k]$ é um vetor de tamanhos de memória.

Outra maneira de melhorar o desempenho do DTW é concatenar o vetor da Eq. 3.4 para várias regras de caminhada (35):

$$\Psi_{\vec{\mu}}^{\vec{r}} = [\psi_{\vec{\mu}}^{r_1}, \psi_{\vec{\mu}}^{r_2}, \dots, \psi_{\vec{\mu}}^{r_j}] \quad (3.5)$$

onde $\vec{r} = [r_1, r_2, \dots, r_j]$ é o vetor de regras de caminhada.

O algoritmo possui três condições de parada: a primeira delas é caso o turista encontre um atrator e, desta forma, o algoritmo retorna transiente $t = \tau$ e atrator $a = \alpha$. A segunda condição ocorre quando o turista encontra uma cidade da extremidade do mapa, *i.e.*, o único vizinho é a cidade do passo anterior e que, por restrição de memória, o turista não pode retornar a ela. Neste caso o turista para e retorna transiente $t = \tau$ e atrator $a = 0$. Por último, se o turista exceder uma quantidade máxima e pré-definida de passos (tipicamente valendo $\mu + z$), então o algoritmo retorna transiente $t = \mu + z$ e atrator $a = 0$.

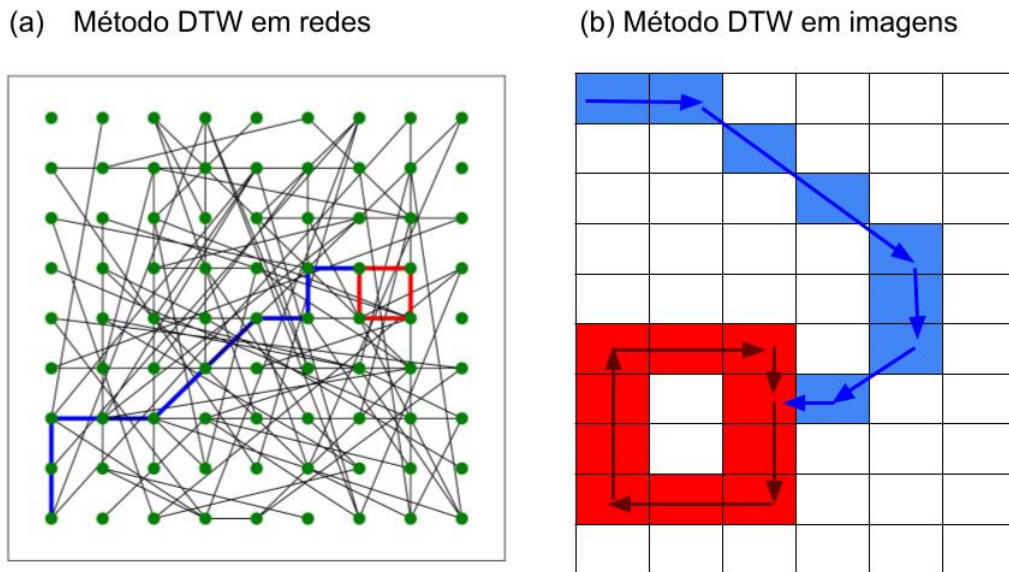


Figura 8 – Exemplo de trajetória de uma caminhada determinística do turista em: (a) Redes. (b) Imagens. As setas em azul indicam a parte transitiente da trajetória e as setas em vermelho representam o atrator.

Fonte: Elaborada pelo autor.

3.1 O método DTW aplicado à análise de texturas

A análise de texturas é a área para a qual o algoritmo da caminhada determinística do turista foi originalmente desenvolvido. (8,33–35) Usando a analogia do turista, a imagem é o mapa e os pixels da imagem são as cidades. Então, dada a matriz M que representa todos os pixels da imagem, o turista se desloca do pixel i ao pixel j seguindo duas regras possíveis:

1. **T-R1:** O turista caminha em direção ao pixel j que minimiza a diferença de intensidade de cor (ou nível de cinza) dI em relação ao pixel atual.
2. **T-R2:** O turista caminha em direção ao pixel j que maximiza a diferença de intensidade de cor (ou nível de cinza) em relação ao pixel atual.

onde $dI = |I_i - I_j|$ é a diferença da intensidade de cor ou de nível de cinza (I) entre os pixels i e j .

Uma vez executado o algoritmo DTW partindo de cada pixel da imagem, então os vetores das Eq. 3.3, 3.4 e 3.5 podem ser obtidos para cada imagem, o que permitirá a classificação delas por um método de ML.

Devido ao alto custo computacional do método, o que inviabilizaria o método para imagens com alta resolução, uma alternativa foi estudada para o DTW, onde ao invés de lançar o turista de cada pixel da imagem, um conjunto menor de pixels é selecionado e essa seleção deve ser também determinística (*e.g.*, apenas os pixels rotulados por número par são selecionados). Mesmo com a redução da quantidade de trajetórias, o método apresentou bons resultados. (36)

A forma tradicional de executar o algoritmo DTW em imagens é através da matriz de pixels, como mostra a Figura 8.(b). No entanto, outras formas foram desenvolvidas as quais unem análise de imagens com redes complexas. Uma imagem pode ser convertida em rede, onde cada pixel torna-se um nó, mantendo a propriedade da intensidade de cor. O peso de cada aresta é equivalente à diferença de intensidade de cor dI entre vizinhos. Dessa forma, a rede gerada conterá todos os caminhos produzidos pelo turista na imagem seguindo uma das duas regras estabelecidas. (35,37) A rede será única para cada imagem e medidas podem ser extraídas para construir o vetor de características. Tanto o histograma conjunto de transiente e atrator quanto o histograma de grau da rede podem ser usados como descritores.

A Figura 9 mostra, na esquerda, a imagem original exibida em termos dos níveis de cinza e a imagem da direita foi produzida pelo turista através de uma rede complexa e cada pixel assume o valor de grau do correspondente vértice. Portanto, neste último caso o método DTW atuou como um filtro, realizando operações matemáticas sobre a imagem e convertendo níveis de cinza em graus de uma rede correspondente.

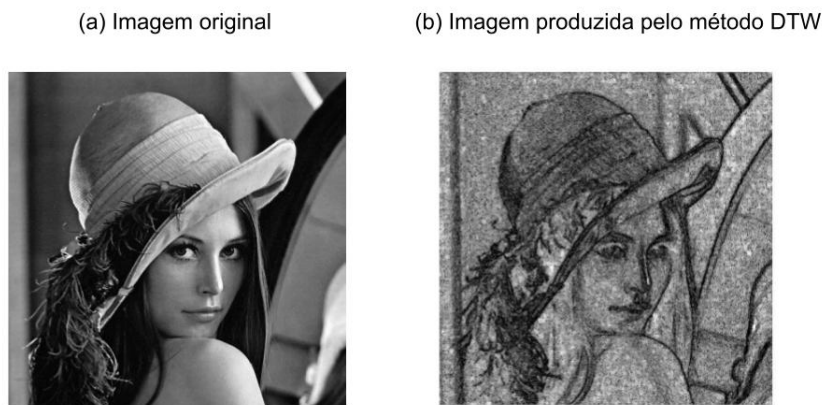


Figura 9 – (a) Imagem original. (b) Imagem produzida através de uma rede complexa modelada pelo algoritmo DTW. Os tons de cinza equivalem ao valor de grau do vértice correspondente ao pixel. Para uma maior resolução, os valores de grau foram convertidos numa escala de 0 à 255 usando uma transformação logarítmica.

Fonte: (a) The USC-SIPI image database. (38) (b) Elaborada pelo autor.

3.2 O método DTW aplicado à redes

O uso da caminhada determinística do turista diretamente em redes, como mostra a Figura 8.(a), foi apresentado por *Gonçalves et. al.* (39), onde o algoritmo do turista foi executado em quatro modelos de rede (Redes geométricas, redes livre de escala, redes de pequeno-mundo e redes aleatórias) extraindo de cada rede os vetores de característica definidos pelas Equações 3.3, 3.4 e 3.5 e então classificando as redes. O estudo também analisou o comportamento, do ponto de vista de medidas estatísticas do turista, em cada um desses modelos.

No artigo de *Gonçalves et. al.* (39) foram definidas duas regras de caminhada (regras N-R1 e N-R2) baseadas na distância entre vértices, mas podemos definir outras duas (N-R3 e N-R4) baseadas na diferença de grau. As regras são, portanto, as seguintes:

1. **N-R1:** O turista caminha em direção ao vértice mais próximo.
2. **N-R2:** O turista caminha em direção ao vértice mais afastado.
3. **N-R3:** O turista move-se para o vértice que minimiza a diferença de grau.
4. **N-R4:** O turista move-se para o vértice que maximiza a diferença de grau.

Em outras palavras, as regras N-R1 e N-R2 buscam minimizar ou maximizar, respectivamente, a distância entre vértices e as regras N-R3 e R-N4 buscam otimizar a diferença de grau. Outras regras, otimizando outras medidas, como o coeficiente de aglomeração, podem ser usadas para movimentar o turista pela rede.

3.3 O método DTW aplicado à redes de pequeno-mundo (DTW-SW)

Uma variação do método da caminhada determinística do turista foi proposta pelo presente projeto e desenvolvida entre os anos de 2021 e 2022 para explorar a transição entre rede regular e rede de pequeno-mundo e a transição entre rede de pequeno-mundo e rede aleatória em grafos de Watts-Strogatz. Para diferenciar do método original, ele será denotado por DTW-SW. O método possui duas mudanças em relação aos métodos anteriores. A primeira mudança foi realizada nas regras de caminhada. Os algoritmos anteriores apresentaram duas regras, uma que maximiza a diferença (de grau ou de intensidade de cor) e outra que minimiza essa diferença. Ambas são independentes e o turista usa uma ou outra, mas não as duas simultaneamente. Já nessa nova versão do DTW, há duas regras também, mas elas são dependentes uma da outra *i.e.*, a primeira regra deve ser satisfeita para então o turista estar sujeito a segunda regra. As regras são as seguintes:

1. **SW-R1:** O turista estando no vértice i caminhará para o vértice j se e somente se o grau do vértice j diferir do grau do vértice i e este vértice não estiver em sua memória.
2. **SW-R2:** O turista seleciona dentre os vizinhos que satisfazem a regra SW-R1, aquele que minimiza a diferença de coeficiente de aglomeração.

O turista está, portanto, sujeito tanto à distribuição de grau pela regra SW-R1 quanto à distribuição de coeficiente de aglomeração pela regra SW-R2, sendo que o coeficiente de aglomeração é uma medida importante para redes de pequeno-mundo, como comentado na Seção 2.2.3

Note que, se o turista estiver em um nó onde todos os vizinhos possuem o mesmo grau, então ele permanece parado, encerrando assim o movimento. Isso implica que para redes regulares o turista terá trajetória nula para qualquer vértice inicial e para redes parcialmente regulares, terá baixa média de comprimento das trajetórias.

A segunda modificação foi no vetor de característica descrito pela Eq. 3.3. Foi adicionado o termo $h(\ell = 0)$ que está associado a trajetórias nulas, sendo esse termo, portanto, um indicador de regularidade da rede, onde se a rede for regular $h(\ell = 0) \rightarrow 1$ e se a rede for aleatória $h(\ell = 0) \rightarrow 0$. O novo vetor de característica torna-se, portanto:

$$\phi_\mu = [h(\ell = 0), h(\ell = \mu + 1), h(\ell = \mu + 2), h(\ell = \mu + 3), \dots, h(\ell = \mu + z)] \quad (3.6)$$

Os vetores definidos pelas Eq. 3.4 e Eq. 3.5 permanecem sendo definidos da mesma forma.

Esse método possui quatro condições de paradas para cada trajetória:

1. **Rede regular:** Se a rede for regular, o turista não caminha, retornando transiente $t = 0$ e atrator $a = 0$.
2. **Rede localmente regular:** Se o turista encontrar um vértice da rede o qual todos os vizinhos possuem o mesmo grau *i.e.*, a rede é localmente regular (ou uniforme), então o turista encerra a trajetória retornando $t = \tau$ e $a = 0$.
3. **Encontrou um atrator:** Nesse caso o turista retorna $t = \tau$ e $a = \alpha$ como comprimento do transiente e atrator, respectivamente.
4. **Exceder o limite de passos:** Se o turista exceder o limite de passos $\mu + z$, então a trajetória termina retornando $t = \mu + z$ e $a = 0$

O parâmetro z foi definido como sendo 500, um valor alto o suficiente para que somente em casos excepcionais o turista fique na iminência de exceder esse valor. Cada passo tem o comprimento igual a 1.

A partir desse método foi proposto uma nova métrica para medir o quanto uma rede se assemelha a redes de pequeno-mundo e que, à semelhança dos índices σ e ω , definidos pelas Eq. 2.32 e Eq. 2.33, usam uma propriedade topológica da rede (coeficiente de aglomeração), mas diferente deles essa nova métrica, denotada por χ , usa também o comprimento médio da trajetória do turista na rede normalizada pelo comprimento médio da trajetória do turista em uma rede aleatória equivalente. Esse coeficiente é definido como:

$$\chi = C \cdot \frac{\langle \ell \rangle}{\langle \ell_{random} \rangle} \quad (3.7)$$

onde C é o coeficiente de aglomeração da rede. Espera-se que a trajetória do turista se anule para redes regulares e o coeficiente de aglomeração tenda a zero para redes aleatórias, desta forma, o coeficiente χ só apresentaria valores positivos para redes semelhantes a uma rede de pequeno-mundo.

Como se trata de resultados recentes, importantes, que fazem parte do presente projeto de mestrado, eles serão mostrados e discutidos na Seção 6.5. Os resultados desse método foram publicados pela revista *Physica A: Statistical Mechanics and its Applications*.
(40)

4 A CAMINHADA DETERMINÍSTICA DO TURISTA COM BIFURCAÇÕES

Considere o algoritmo DTW da Seção 3.2 caminhando em uma rede conforme a regra N-R3. Suponha que em um dado momento o turista chegue a um vértice e aplique a citada regra para continuar seu movimento, como mostram os vetores abaixo:

$$\begin{cases} U_i = [55, 56, 104, 195, 45, 68, 24] \\ d_i = [4, 28, 1, 3, 1, 2, 6] \\ \min(d_i) = 1 \rightarrow \text{vértices 104 e 45} \end{cases} \quad (4.1)$$

onde U_i é o vetor de vizinhos do vértice i e d_i o vetor de diferença de grau associado ao vetor U_i .

Entretanto, há um problema. Dois vértices vizinhos satisfazem a regra N-R3 (os vértices 104 e 45). Esses dois vizinhos possuem o mesmo valor de diferença de grau e esse valor é justamente o mínimo, aquele que deveria ser escolhido pelo turista segundo a dada regra. Esse caso pode ocorrer. O turista foi programado para escolher apenas um vértice, não dois ou três. Existem diversas maneiras de resolver tais conflitos no método DTW. Uma das soluções é selecionar, a partir da lista de adjacências U_i do vértice atual i , o primeiro vértice que satisfaça a regra de caminhada, garantindo assim a manutenção do determinismo no processo. Neste exemplo, o vértice 104 foi selecionado, enquanto que o vértice 45 foi ignorado pelo turista. Ignorar os outros vértices que empataram implica em perda de informação e para evitar essa perda foi desenvolvido um novo autômato baseado no DTW.

Esse novo autômato possui uma regra fundamental rotulada como **B-R0**, por ser uma regra que não pode ser escolhida e sim uma regra inerente ao algoritmo. A regra B-R0 diz que a cada empate um clone do turista é criado possuindo o mesmo tamanho de memória e, por consequência, os mesmos vértices armazenados na memória do turista original estão também armazenados na memória do clone. O clone também possui o mesmo tamanho de transiente (considerando até o nó onde ocorreu a clonagem). São de fato clones e não é possível distinguir um clone de seu original. Cada clone seguirá em direção a um dos vértices onde houve empate, criando assim, bifurcações na trajetória, sendo esse novo autômato, portanto, chamado de “caminhada determinística do turista com bifurcações” ou simplesmente “DTW com bifurcações” (DTWB). A Figura 10 ilustra uma trajetória com bifurcações.

Retomando o exemplo do início do capítulo, há dois empates no vértice i , então o turista gera um clone de si e enquanto um segue em direção ao vértice 104 o outro segue rumo ao vértice 45. Se houver três empates, então dois clones serão criados e assim por

diante.

Uma vez definido a regra $B - R0$, podemos definir as regras de caminhada, são elas:

1. **B-R1:** O turista caminha em direção ao vértice que possui o mesmo grau k .
2. **B-R2:** O turista segue em direção ao vértice que possui grau k diferente em relação ao vértice atual.
3. **B-R3:** O turista move-se para o vértice que minimiza a diferença de grau.
4. **B-R4:** O turista move-se para o vértice que maximiza a diferença de grau.

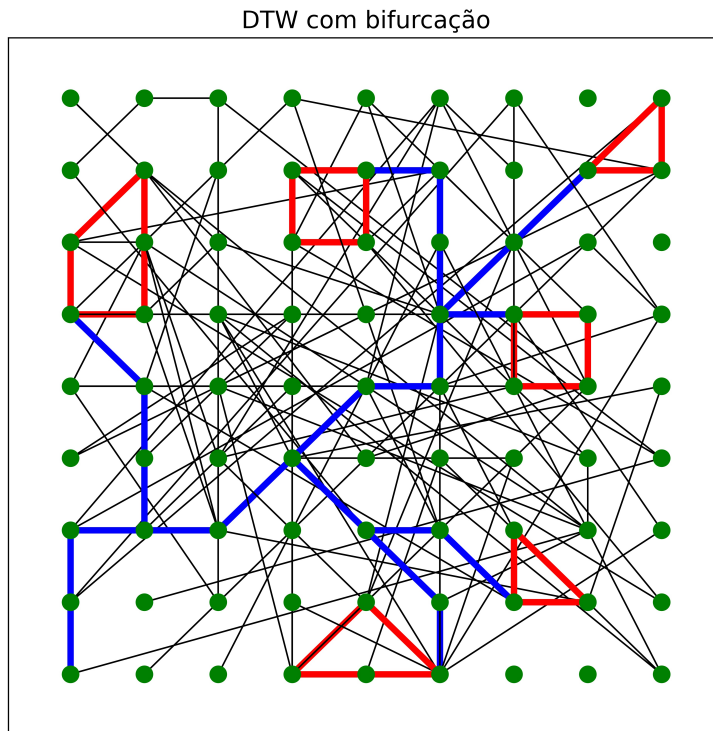


Figura 10 – Caminhada determinística do turista com bifurcações (DTWB)

Fonte: Elaborada pelo autor.

Podemos então definir uma nova medida para a rede, o grau (ou nível) de bifurcação, que quantifica quantas bifurcações surgem a partir de um vértice, como mostra a Figura 11. O turista marcará cada nó pelo qual ele passar com esse valor. O turista será lançado de cada nó da rede, então cada nó terá um grau de bifurcação.

O Grau de bifurcação pode ser numericamente definido como a quantidade de empates menos a quantidade de nós empatados que estão na memória do turista. Para memória $\mu \geq 1$ e para nós que não são um nó inicial da trajetória, o grau de bifurcação possui valor mínimo (excluindo o caso de B nulo) de $B = qtd. de empates - 1$, pois devido ao tamanho de memória, o nó antecessor está incluindo na memória e o mesmo satisfaz a regra em questão, então apesar de o nó antecessor ser um empate, não haverá bifurcação em direção a ele. Seja X_i^r o conjunto de vizinhos empatados no i -ésimo vértice de acordo com regra r e Y_i^μ , o conjunto de vizinhos de i armazenados na memória do turista, então B é definido como:

$$B = |X_i^r| - |X_i^r \cap Y_i^\mu| \quad (4.2)$$

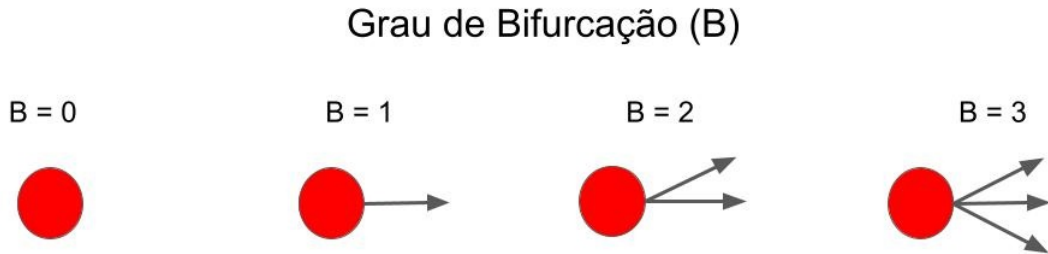


Figura 11 – Grau de bifurcação

Fonte: Elaborada pelo autor.

Outra medida pode ser realizada, a razão entre o grau de bifurcação (B) e a quantidade de vizinhos do nó. Em outras palavras, a medida β é a razão entre o grau de bifurcação e o grau topológico.

$$\beta = \frac{B}{\text{quantidade de vizinhos}} = \frac{B}{k} \quad (4.3)$$

A taxa β , é a densidade de B e mede quantos empates há por número de vizinhos. Para redes dirigidas, o índice β é a razão da quantidade de bifurcações pela quantidade de nós sucessores, *i.e.*, a razão do grau B sobre o grau de saída k_{out} . Caso a rede tenha pesos, o grau topológico é substituído pela quantidade de vizinhos.

Sabendo que as regras B-R1 e B-R2 são exatamente opostas e sabendo haver apenas duas opções no que se refere ao grau k ; ou os vizinhos do i -ésimo nó possuem grau diferente do nó i , ou possuem o mesmo grau, então segue a seguinte propriedade para β :

$$\beta_{i,B-R1} + \beta_{i,B-R2} \approx 1 \quad (4.4)$$

A soma das taxas β para as regras B-R1 e B-R2, para um dado nó i , tende a um. Devido à influência da memória no valor do grau de bifurcação, a Equação 4.4 não é identicamente igual a um, mas um valor aproximado. A Equação 4.4 torna-se identicamente igual a um, caso a memória seja nula. Dada a linearidade da média, podemos expandir tal propriedade para o β global da rede:

$$\frac{1}{n} \sum_{i=1}^n \beta_{i,B-R1} + \frac{1}{n} \sum_{i=1}^n \beta_{i,B-R2} = \frac{1}{n} \sum_{i=1}^n (\beta_{i,B-R1} + \beta_{i,B-R2}) \approx 1 \quad (4.5)$$

O grau k é o grau topológico da rede definido na Seção 2.1. Uma das quatro regras de movimento é selecionada e se houver empates de acordo ela, então haverá bifurcações na trajetória.

A trajetória do turista permanece sendo formada por duas partes, uma parte transiente e o atrator (Ver Figura 10), com a diferença que devido às bifurcações, haverá mais trajetórias, explorando outras regiões da rede e extraindo mais informações do sistema.

4.1 O vetor de característica

Foram tomadas quatro medidas para o vetor de característica:

1. Grau de bifurcação (B)
2. Bifurcações/vizinhos (β)
3. comprimento do transiente (t)
4. comprimento do atrator (a)

Cada nó da rede terá um grau de bifurcação e um índice β e cada trajetória produzida pelo turista terá um transiente e um atrator, mesmo que estes sejam nulos. Para compor o vetor de característica da rede foram tomadas três medidas estatísticas de cada uma das quatro medidas citadas acima: média ($\langle \rangle$), desvio padrão (σ) e entropia de Shannon (H). O vetor é definido como:

$$\phi_{\mu}^r = [\langle B \rangle, \sigma_B, H_B, \langle \beta \rangle, \sigma_{\beta}, H_{\beta}, \langle t \rangle, \sigma_t, H_t, \langle a \rangle, \sigma_a, H_a] \quad (4.6)$$

Outras medidas foram testadas para o vetor de característica, incluindo o histograma de grau de bifurcação e o histograma de densidade de bifurcação (β) e os resultados destes vetores alternativos serão apresentados na Seção 6.3, no entanto, o vetor descrito pela Eq. 4.6 apresentou o melhor resultados e é através dele que será executado o experimento de classificação de redes cujos resultados estão na Seção 6.4.

4.2 O método DTWB é determinístico?

Uma das perguntas mais importantes que se faz após a definição do método DTWB é se ele mantém o determinismo de seu antecessor. A resposta é sim. A criação dos autômatos clones se dá de forma plenamente determinística, pelos empates a partir da aplicação das regras de caminhada, as quais são determinísticas. Como a trajetória do turista deriva dessas regras, por consequência, as trajetórias também são determinísticas.

5 MÉTODO EXPERIMENTAL

5.1 Datasets

Durante os experimentos foram utilizados cinco conjuntos de redes:

1. **Dataset 1:** Um *dataset* composto por redes de quatro modelos: Erdos-Renyi (com $p = \langle k \rangle / n - 1$), Barabási-Albert (com quantidade de novas conexões por nó igual à 3), redes de Watts-Strogatz (com $p = 0, 1$) e Waxman (com $\alpha = 0, 1$). Cada modelo terá 28 combinações de tamanho e grau médio e cada uma dessas combinações terá 100 redes, totalizando 11200 redes neste *dataset*. Foram considerados quatro tamanhos: $n = 500, 1000, 1500, 2000$ nós, e sete valores de grau médio $\langle k \rangle = 4, 6, 8, 10, 12, 14, 16$.
2. **Dataset 2:** Esse *dataset* é o mesmo que o **Dataset 1**, porém com a adição de ruído para verificar a robustez do método proposto. São três taxas de ruído: $\rho = 10\%, 20\%, 30\%$, onde a adição de ruído significa eliminar e adicionar aleatoriamente arestas do grafo, sendo assim $\rho/2$ das arestas serão eliminadas e $\rho/2$ arestas serão adicionadas.
3. **Dataset 3:** Esse *dataset* contém redes do mundo real extraídas do *KONECT project*. (24)
4. **Dataset 4:** Redes metabólicas extraídas do *Kyoto Encyclopedia of Genes and Genomes* (KEEG) (41), onde cada vértice representa um metabólito e as arestas representam a conexão entre metabólitos e substratos. Esse conjunto é dividido em 7 subconjuntos:
 - **Actinobacteria:** Consiste em três classes: *Corynebacterium*, *Mycobacterium* e *Streptomyces* com 86, 60 e 53 espécies, respectivamente.
 - **Animais:** Consiste em quatro classes, cada classe com 14 espécies, são elas: insetos, mamíferos, pássaros e peixes.
 - **Firmicutes-Bacillus:** Consiste em quatro classes: *Bacillus*, *Lactobacillus*, *Staphylococcus* e *Streptococcus*, com 128, 83, 76 e 133 espécies, respectivamente.
 - **Fungi:** Esse *dataset* contém quatro classes: *Basid-iomycetes*, *eurotiomycetes*, *saccharomycetes* e *sordariomycetes*, com 15 espécies em cada classe.
 - **Plantas:** São três classes: *Eudicots*, *Green Algae* e *Monocots*, com 9 espécies em cada uma.
 - **Protistas:** São quatro classes: *Amoebozoa*, *Alveolates*, *Euglenozoa* e *Stramenopiles*, cada classe possui 5 organismos.

– **Reinos:** São quatro classes, com quarenta redes em cada uma delas. As classes são: *Animais, Fungos, Plantas e Protistas*.

5. **Dataset 5:** É composto por 100 redes sociais com duas classes: *Google+* e *Twitter*, cada classe com 50 redes. Essas redes foram extraídas do *Stanford Network Analysis Project* (SNAP). (42)

5.2 Métodos de classificação e validação

Nos experimentos foram utilizados dois classificadores, o *Linear discriminant analysis* (LDA) e o *Principal component analysis* (PCA).

Para a validação foi utilizado o método *Leave-one-out* (LOOCV). O LOOCV é um método de validação cruzada que avalia a capacidade de predição do classificador. Os métodos classificadores LDA, PCA e a abordagem LOOCV para a validação serão apresentados formalmente no Apêndice A.

5.3 Métodos de comparação

Foram utilizados três métodos para a comparação de desempenho do método proposto na Seção 4, são eles:

1. Caminhada determinística do turista (tradicional): O método apresentado na Seção 3.2 foi usado na comparação. Os vetores de característica utilizados foram os vetores definidos pela Equação 3.4 e Equação 3.5. Foi utilizado a seguinte combinação de memória: $\vec{\mu} = [1, 2, 3, 4]$ e duas regras de caminhada: N-R3 e N-R4. Em suma, serão três vetores de característica para cada rede: $\psi_{[1, 2, 3, 4]}^{N-R3}$, $\psi_{[1, 2, 3, 4]}^{N-R4}$ e $\Psi_{[1, 2, 3, 4]}^{[N-R3, N-R4]}$.
2. LLNA, discutido na Seção 2.3.2. O vetor de característica foi construído através do padrão binário, como descreve o artigo de Ribas *et al.* (31), inicializando a rede com 50% das células vivas e executando a simulação por 350 iterações. As regras escolhidas foram: B135678-S03456 para as redes sintéticas (*dataset 1* e *dataset 2*) e a regra B02345678-S123468 para as redes reais.
3. Medidas estruturais. Foram tomadas as seguintes medidas: Grau médio ($\langle k \rangle$), desvio padrão do grau (σ_k), coeficiente de aglomeração (C), média do comprimento dos menores caminhos ($\langle l \rangle$) e a assortatividade de grau usando a correlação de Pearson (r_r). Sendo assim, o vetor de característica é definido como:

$$\Lambda = [\langle k \rangle, \sigma_k, C, l, r_r] \quad (5.1)$$

5.4 Comparação de tempo computacional

Para a comparação do tempo de execução dos algoritmos, foi utilizada, em todos os experimentos, a mesma configuração computacional:

1. CPU: AMD Ryzen Threadripper 1950X, 32 cores, 3.3 GHz
2. Memória RAM: 96 GB
3. GPU: RTX 2080 ti, com 12 GB de memória
4. Sistema: 64-bits

6 RESULTADOS E DISCUSSÕES

Neste capítulo serão apresentados os principais resultados e a discussão acerca deles.

6.1 Complexidade computacional

O ponto negativo do método proposto é a sua complexidade computacional. O DTW tradicional é lançado de cada nó da rede e realiza, a partir de cada lançamento, uma trajetória de tamanho $\ell = t + a$. Então, a complexidade do DTW tradicional é da ordem de $\mathcal{O}(n(t + a))$, onde n é o tamanho da rede. Se ℓ for muito menor do que n , então a complexidade é equivalente a $\mathcal{O}(n)$. Somente uma trajetória é permitida por lançamento. Entretanto, no DTWB cada lançamento poderá resultar em mais de uma trajetória devido às bifurcações. Outra implicação do método é que a cada nova bifurcação da trajetória surgem outras bifurcações, aumentando, portanto, a quantidade de trajetórias em progressão geométrica. Suponha o caso excepcional onde cada nó da rede tenha um grau de bifurcação igual a dois. Na primeira iteração, partindo de um nó, duas trajetórias são geradas. Na segunda iteração cada trajetória gera mais duas trajetórias, então serão quatro novas trajetórias. Na terceira iteração serão oito e na quarta iteração dezesseis. Então a complexidade computacional do algoritmo DTWB é, no caso médio, $\mathcal{O}(n(t + a)\bar{B}^{it})$, onde \bar{B} é o grau de bifurcação médio da rede e it a quantidade de iterações. No melhor dos casos (para B não nulo), onde $B = 1$ para todos os nós, retomamos a complexidade do DTW tradicional. Para a regra B-R1 o pior caso é uma rede regular, onde todos os vizinhos possuem o mesmo grau k , portanto, haverá um número elevado de bifurcações por nó. Já para a regra B-R2, o pior caso ocorre quando a rede é bastante heterogênea, como as redes livres de escala e de Waxman. A quantidade de trajetórias aumenta, portanto, exponencialmente com o número de iterações, como mostra a Figura 12.

Há duas saídas possíveis para este problema. A primeira é limitar a quantidade de iterações. Entretanto, se aplicarmos essa solução, deveríamos limitar a quantidade de iterações a não mais do que 10 ou 15 iterações e, a depender da rede, muitas trajetórias ainda não terão encontrado um atrator. A segunda solução, mais segura, é limitar a quantidade de bifurcações e, após atingir esse limite, permitir que cada trajetória gerada seja concluída de forma semelhante ao DTW tradicional. O grau de bifurcação continuará a ser computado, mas não mais bifurcações serão geradas e somente o primeiro nó do empate será escolhido, a exemplo do DTW original. O limite escolhido para todos os experimentos realizados foi de 1000 bifurcações, um valor alto o suficiente para permitir a extração de padrões da rede, mas não tão alto para ser demasiadamente custoso computacionalmente.

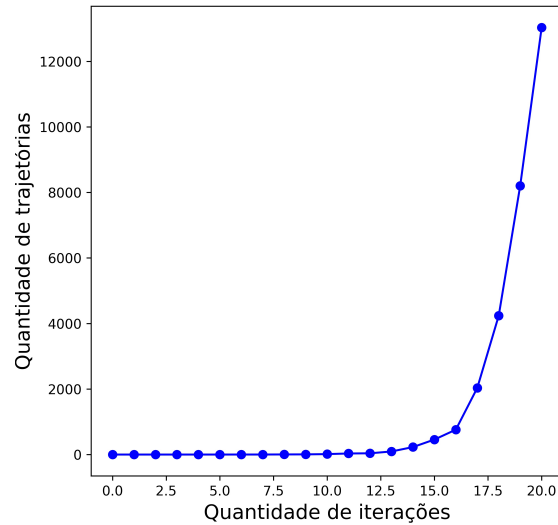


Figura 12 – Quantidade de trajetórias geradas, partindo de um único nó, a cada iteração. Rede: Watts-Strogatz, com 500 nós e grau médio igual a seis. O turista possui memória $\mu = 2$.

Fonte: Elaborada pelo autor.

6.2 Análise do método DTWB

Esta seção contém a análise estatística do método DTWB no que concerne as medidas que compõem o vetor de característica (grau de bifurcação, densidade de bifurcação, transiente e atrator) para cada uma das quatro regras de caminhada definidas no Capítulo 4. A análise será feita nos quatro principais modelos de redes: redes aleatórias (ER), redes de pequeno-mundo (WS), redes livres de escala (BA) e redes geométricas (Waxman).

6.2.1 Grau de bifurcação (B) e densidade de bifurcação (β)

A regra B-R1 favorece redes com alta uniformidade de grau como as redes regulares e redes de Watts-Strogatz com probabilidade de religação baixa. As Tabelas 1 e 2 mostram que a maior média de grau de bifurcação ocorreu para redes de Watts-Strogatz, bem como o modelo WS também obteve os maiores valores de β dentre os modelos analisados. Como esperado, os menores valores de grau de bifurcação, bem como de sua densidade β , ocorrem para as redes de Barabasi-Albert, seguido pelas redes de Waxman e Erdos-Renyi. A Figura 13 mostra os graus de bifurcação de cada um dos modelos.

A regra B-R2 é o oposto da regra B-R1, portanto, espera-se que as distribuições de grau de bifurcação demonstrem essa oposição.

As Tabelas 3 e 4 mostram que as redes livres de escala e as redes de Waxman possuem as maiores médias de grau de bifurcação para a regra B-R2, a qual favorece

redes pouco uniformes. Como esperado, as redes WS possuem a menor média de B . As redes BA, ER e Waxman possuem alta taxa β , indicando, portanto, que quase todos os vizinhos possuem um grau diferente e como consequência, gerando uma alta incidência de empates, *i.e.*, bifurcações. As redes WS mantiveram o valor de β próximo ao valor obtido para a regra B-R1. A Figura 14 mostra a distribuição de grau de bifurcação para a regra B-R2. A distribuição de B tende a Poisson para as redes de Waxman, Erdos-Renyi e Watts-Strogatz, enquanto as redes de Barabasi-Albert possuem distribuição de B em forma de lei de potência. Para essa regra, a distribuição de grau de bifurcação resultou em uma distribuição semelhante a de grau topológico sugerindo uma alta correlação entre as duas métricas para esta regra de caminhada, fato também evidenciado pelo valor de β , o qual é próximo de 1 para três dos quatro modelos, a exceção são a redes do modelo WS.

Tabela 1 – Média e desvio padrão do grau de bifurcação B . Regra: B-R1

Modelo	$B(\mu = 1)$	$B(\mu = 2)$	$B(\mu = 3)$	$B(\mu = 4)$
Barabasi-Albert	$0,63 \pm 0,68$	$0,37 \pm 0,59$	$0,37 \pm 0,59$	$0,26 \pm 0,49$
Erdos-Renyi	$1,25 \pm 0,92$	$1,13 \pm 1,02$	$0,94 \pm 0,96$	$0,98 \pm 0,92$
Waxman	$1,68 \pm 1,10$	$1,44 \pm 1,47$	$1,18 \pm 1,03$	$1,57 \pm 1,29$
Watt-Strogatz	$3,36 \pm 1,18$	$2,84 \pm 1,81$	$2,69 \pm 1,42$	$2,88 \pm 1,49$

Fonte: Elaborada pelo autor.

Tabela 2 – Média e desvio padrão da razão β . Regra: B-R1.

Modelo	$\beta(\mu = 1)$	$\beta(\mu = 2)$	$\beta(\mu = 3)$	$\beta(\mu = 4)$
Barabasi-Albert	$0,11 \pm 0,13$	$0,06 \pm 0,05$	$0,08 \pm 0,06$	$0,05 \pm 0,10$
Erdos-Renyi	$0,18 \pm 0,13$	$0,18 \pm 0,15$	$0,17 \pm 0,15$	$0,13 \pm 0,13$
Waxman	$0,10 \pm 0,06$	$0,08 \pm 0,06$	$0,06 \pm 0,05$	$0,08 \pm 0,07$
Watt-Strogatz	$0,55 \pm 0,19$	$0,47 \pm 0,21$	$0,44 \pm 0,23$	$0,47 \pm 0,25$

Fonte: Elaborada pelo autor.

Diferentemente das regras B-R1 e B-R2, as regras B-R3 e B-R4 não são complementares. Como as duas últimas lidam com a diferença de grau, pode haver vizinhos cuja a diferença de grau não é a máxima e nem a mínima. Já as regras B-R1 e B-R2 lidam com os valores absolutos de grau e nesse caso há somente duas opções, ou o vizinho possui o mesmo grau que o dado vértice ou um grau diferente dele, fazendo dessas duas regras complementares. Então, espera-se que não haja a mesma variação em B e β entre as regras B-R3 e B-R4 como houve entre as regras B-R1 e B-R2. As Tabelas 5 e 6 mostram os valores de B e de β , respectivamente, para a regra B-R3. As distribuições de B para a regra B-R3 são mostradas na Figura 15. As Tabelas 7 e 8 mostram as médias de B e β ,

respectivamente, para as regras B-R4. A Figura 16 mostra a distribuição de B para a regra B-R4. Podemos notar pelas citadas tabelas que os valores de B e β são, em geral, baixos para as duas regras, exceto para as redes WS que apresentou os maiores valores de B e β para ambas as regras. Houve um aumento da média de β para as redes BA na regra B-R4 com relação à regra B-R3 e há uma queda na média de β para as redes WS para a regra B-R4 com relação à regra B-R3. O primeiro caso pode ser explicado pela estrutura da rede. Dado o processo de ligação preferencial, os vértices com baixo grau, os quais são a maioria nesta rede, tendem a se conectar com *hubs*, que possuem alto grau, maximizando, portanto, a diferença de grau, enquanto os *hubs* tendem a se conectar com outros *hubs*, mas estes são poucos na rede. Desse modo, há muito mais conexão que maximizam a diferença de grau entre os vértices ligantes do que conexões que minimizam essa diferença. Já no caso das redes WS, dado o tipo de distribuição de grau (de Poisson) e a baixa variância, essa rede favorece a minimização da diferença de grau entre dois vértices.

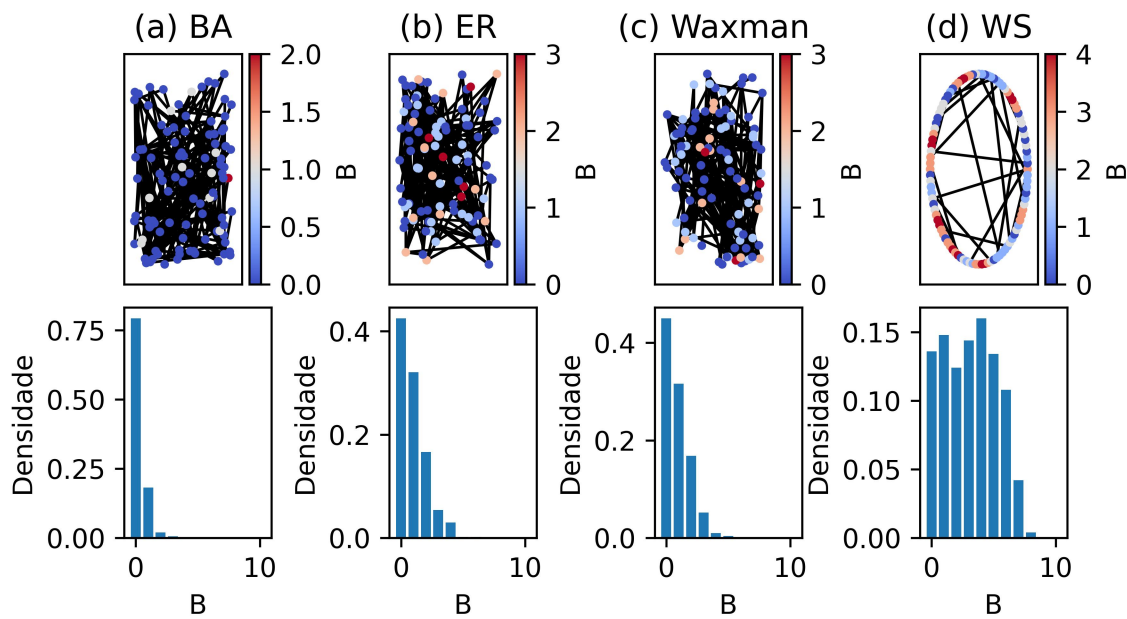


Figura 13 – Os quatro principais modelos de redes (linha de cima) e suas respectivas distribuições de grau de bifurcação B (linha de baixo). (a) Rede BA. (b) Rede ER. (c) Rede de Waxman. (d) Rede WS. Regra de caminhada: B-R1 e memória $\mu = 1$. Cada rede possui 500 nós e grau topológico médio igual a seis.

Fonte: Elaborada pelo autor.

Dada a pouca variação das métricas grau de bifurcação e densidade de bifurcação para as regras de caminhada B-R3 e B-R4, essas métricas não são tão relevantes para tais regras e as medidas de transiente e atrator (ver Seção 6.2.2) mostram uma diferença maior entre modelos para as regras B-R3 e B-R4.

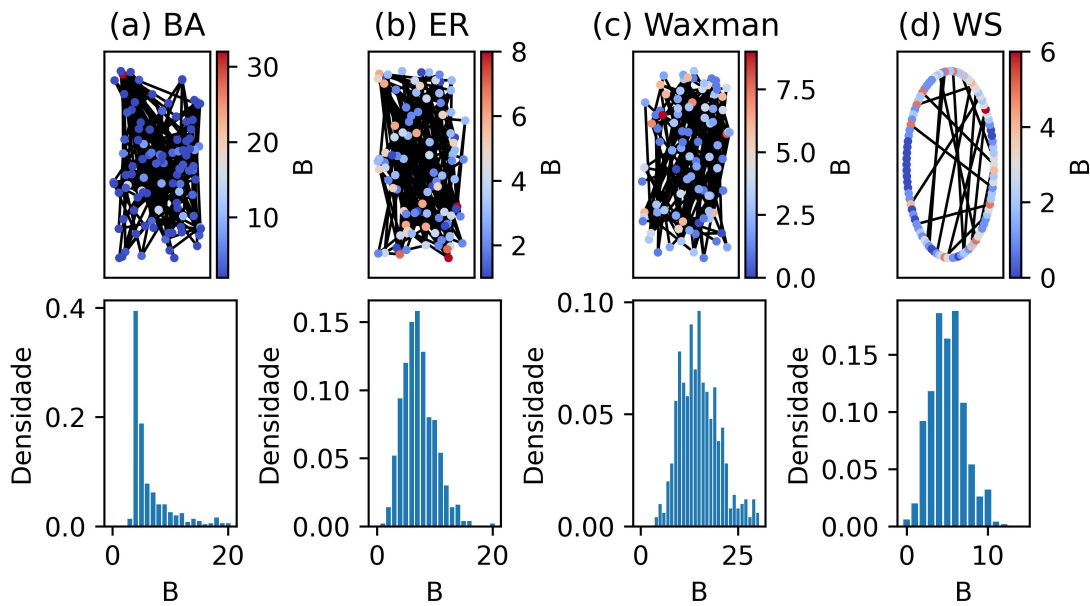


Figura 14 – Os quatro principais modelos de redes (linha de cima) e suas respectivas distribuições de grau de bifurcação B (linha de baixo). (a) Rede BA. (b) Rede ER. (c) Rede de Waxman. (d) Rede WS. Regra de caminhada: B-R2 e memória $\mu = 1$. Cada rede possui 500 nós e grau topológico médio igual a seis.

Fonte: Elaborada pelo autor.

Tabela 3 – Média e desvio padrão do grau de bifurcação B . Regra: B-R2.

Modelo	$B(\mu = 1)$	$B(\mu = 2)$	$B(\mu = 3)$	$B(\mu = 4)$
Barabasi-Albert	$9,78 \pm 11,85$	$9,88 \pm 11,34$	$9,29 \pm 12,55$	$8,21 \pm 10,09$
Erdos-Renyi	$5,84 \pm 2,36$	$6,33 \pm 2,78$	$5,96 \pm 2,70$	$5,90 \pm 2,69$
Waxman	$9,20 \pm 5,94$	$8,17 \pm 3,14$	$7,75 \pm 4,74$	$8,23 \pm 4,18$
Watt-Strogatz	$4,16 \pm 1,65$	$3,38 \pm 1,77$	$3,17 \pm 1,70$	$3,36 \pm 1,78$

Fonte: Elaborada pelo autor.

De modo geral, o grau de bifurcação (B) tende a acompanhar o grau topológico (k), então, B não é uma boa medida para comparar modelos de redes que podem diferir tanto na distribuição de k quanto nos momentos estatísticos do mesmo. A densidade de grau de bifurcação β é uma métrica mais confiável para comparar diferentes modelos e redes, pois consegue captar o quanto uma dada regra de caminhada está favorecendo um modelo em detrimento de outro. Por exemplo, a média de B para as redes de Barabasi-Albert usando a regra B-R2 é bem maior do que a média de B para as redes ER (ver Tabela 3). No entanto, a média da densidade de bifurcação (ver Tabela 4) entre os dois modelos são semelhantes. Então, o grau B deve ser analisado em companhia de sua densidade β .

Tabela 4 – Média e desvio padrão da razão β . Regra: B-R2.

Modelo	$\beta(\mu = 1)$	$\beta(\mu = 2)$	$\beta(\mu = 3)$	$\beta(\mu = 4)$
Barabasi-Albert	$0,91 \pm 0,13$	$0,90 \pm 0,13$	$0,83 \pm 0,17$	$0,85 \pm 0,16$
Erdos-Renyi	$0,86 \pm 0,14$	$0,85 \pm 0,14$	$0,85 \pm 0,14$	$0,85 \pm 0,14$
Waxman	$0,92 \pm 0,07$	$0,85 \pm 0,10$	$0,82 \pm 0,06$	$0,86 \pm 0,08$
Watt-Strogatz	$0,50 \pm 0,30$	$0,54 \pm 0,25$	$0,56 \pm 0,26$	$0,54 \pm 0,26$

Fonte: Elaborada pelo autor.

Tabela 5 – Média e desvio padrão do grau de bifurcação B . Regra: B-R3.

Modelo	$B(\mu = 1)$	$B(\mu = 2)$	$B(\mu = 3)$	$B(\mu = 4)$
Barabasi-Albert	$1,1 \pm 0,34$	$1,20 \pm 0,43$	$1,37 \pm 0,62$	$1,54 \pm 0,70$
Erdos-Renyi	$1,31 \pm 0,70$	$1,88 \pm 0,64$	$1,99 \pm 1,04$	$1,96 \pm 1,21$
Waxman	$1,49 \pm 0,89$	$2,08 \pm 1,25$	$2,17 \pm 1,33$	$2,08 \pm 1,14$
Watt-Strogatz	$2,53 \pm 1,56$	$3,79 \pm 1,5$	$3,69 \pm 1,35$	$3,70 \pm 1,44$

Fonte: Elaborada pelo autor.

Tabela 6 – Média e desvio padrão da razão β . Regra: B-R3.

Modelo	$\beta(\mu = 1)$	$\beta(\mu = 2)$	$\beta(\mu = 3)$	$\beta(\mu = 4)$
Barabasi-Albert	$0,17 \pm 0,08$	$0,20 \pm 0,09$	$0,20 \pm 0,12$	$0,25 \pm 0,15$
Erdos-Renyi	$0,18 \pm 0,09$	$0,18 \pm 0,10$	$0,23 \pm 0,12$	$0,20 \pm 0,14$
Waxman	$0,09 \pm 0,05$	$0,12 \pm 0,07$	$0,13 \pm 0,07$	$0,13 \pm 0,07$
Watt-Strogatz	$0,31 \pm 0,19$	$0,47 \pm 0,08$	$0,45 \pm 0,16$	$0,45 \pm 0,18$

Fonte: Elaborada pelo autor.

Tabela 7 – Média e desvio padrão do grau de bifurcação B . Regra: B-R4.

Modelo	$B(\mu = 1)$	$B(\mu = 2)$	$B(\mu = 3)$	$B(\mu = 4)$
Barabasi-Albert	$1,05 \pm 0,17$	$1,00 \pm 0,00$	$1,02 \pm 0,16$	$1,05 \pm 0,24$
Erdos-Renyi	$1,03 \pm 0,20$	$1,16 \pm 0,40$	$1,46 \pm 0,65$	$1,23 \pm 0,58$
Waxman	$1,02 \pm 0,17$	$1,04 \pm 0,22$	$1,01 \pm 0,22$	$1,08 \pm 0,28$
Watt-Strogatz	$1,56 \pm 1,24$	$2,11 \pm 1,45$	$2,07 \pm 1,56$	$2,67 \pm 2,03$

Fonte: Elaborada pelo autor.

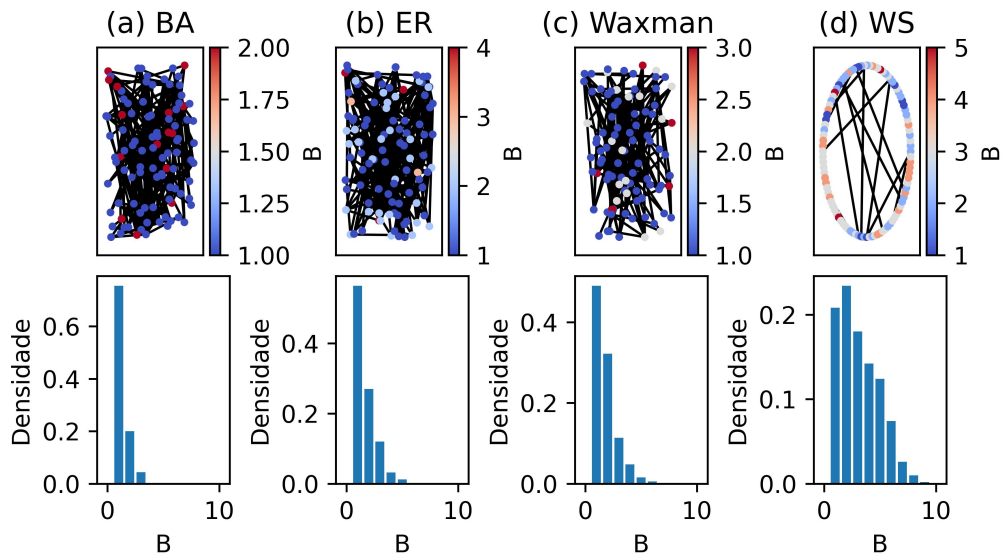


Figura 15 – Os quatro principais modelos de redes (linha de cima) e suas respectivas distribuições de grau de bifurcação B (linha de baixo). (a) Rede BA. (b) Rede ER. (c) Rede de Waxman. (d) Rede WS. Regra de caminhada: B-R3 e memória $\mu = 1$. Cada rede possui 500 nós e grau topológico médio igual a seis.

Fonte: Elaborada pelo autor.

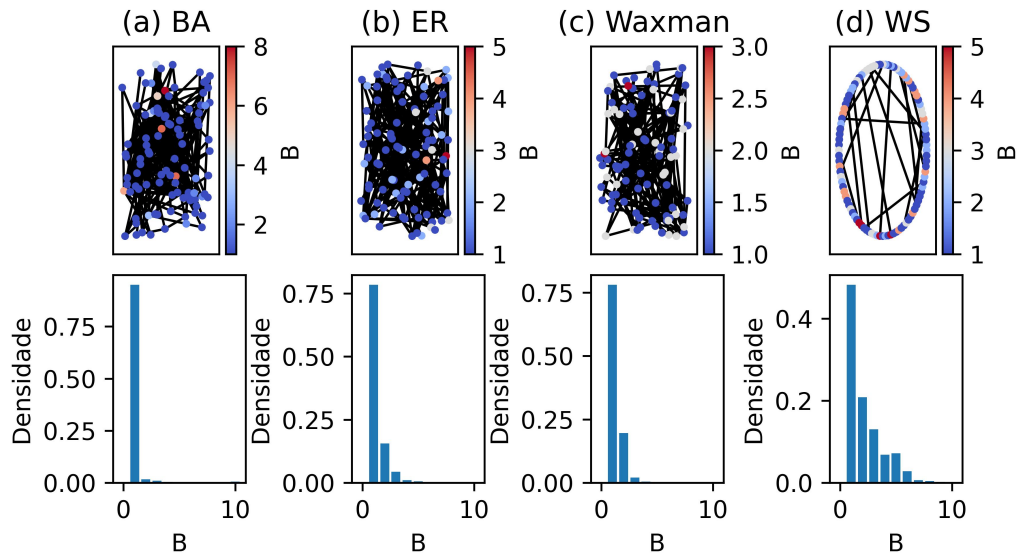


Figura 16 – Os quatro principais modelos de redes (linha de cima) e suas respectivas distribuições de grau de bifurcação B (linha de baixo). (a) Rede BA. (b) Rede ER. (c) Rede de Waxman. (d) Rede WS. Regra de caminhada: B-R4 e memória $\mu = 1$. Cada rede possui 500 nós e grau topológico médio igual a seis.

Fonte: Elaborada pelo autor.

Tabela 8 – Média e desvio padrão da razão β em redes sintéticas. Regra B-R4 e memórias $\mu = 1$, $\mu = 2$, $\mu = 3$ e $\mu = 4$

Modelo	$\beta(\mu = 1)$	$\beta(\mu = 2)$	$\beta(\mu = 3)$	$\beta(\mu = 4)$
Barabasi-Albert	$0,126 \pm 0,05$	$0,25 \pm 0,00$	$0,25 \pm 0,04$	$0,26 \pm 0,06$
Erdos-Renyi	$0,17 \pm 0,12$	$0,20 \pm 0,17$	$0,20 \pm 0,11$	$0,20 \pm 0,11$
Waxman	$0,08 \pm 0,05$	$0,09 \pm 0,05$	$0,09 \pm 0,03$	$0,11 \pm 0,03$
Watt-Strogatz	$0,19 \pm 0,16$	$0,26 \pm 0,17$	$0,26 \pm 0,19$	$0,32 \pm 0,24$

Fonte: Elaborada pelo autor.

6.2.2 Transiente e atrator

As duas últimas medidas presentes no vetor de característica, as quais são herdadas do método DTW tradicional, são o transiente e o atrator. Essas duas medidas dependem fortemente do valor de memória, principalmente o atrator cujo comprimento (não nulo) mínimo é $\mu + 1$. Vale salientar que tanto o transiente quanto o atrator podem ser nulos.

Tabela 9 – Média e desvio padrão dos transientes. Regra: B-R1.

Modelo	$t(\mu = 1)$	$t(\mu = 2)$	$t(\mu = 3)$	$t(\mu = 4)$
Barabasi-Albert	$0,01 \pm 0,1$	$0,01 \pm 0,1$	$0,005 \pm 0,08$	$0,002 \pm 0,03$
Erdos-Renyi	$0,22 \pm 0,54$	$1,68 \pm 4,64$	$1,52 \pm 4,79$	$1,4 \pm 4,7$
Watts-Strogatz	$4,5 \pm 3,7$	$6,2 \pm 7,5$	$7,5 \pm 11,2$	$9,9 \pm 16,1$
Waxman	$0,24 \pm 0,57$	$1,57 \pm 4,17$	$1,54 \pm 4,7$	$1,38 \pm 4,78$

Fonte: Elaborada pelo autor.

Tabela 10 – Média e o desvio padrão dos atratores. Regra: B-R1.

Modelo	$a(\mu = 1)$	$a(\mu = 2)$	$a(\mu = 3)$	$a(\mu = 4)$
Barabasi-Albert	$0,68 \pm 0,92$	$0,07 \pm 0,49$	$0,02 \pm 0,02$	$0,008 \pm 0,02$
Erdos-Renyi	$1,66 \pm 0,72$	$2,02 \pm 2,95$	$1,52 \pm 4,79$	$1,5 \pm 2,2$
Watts-Strogatz	$1,96 \pm 0,28$	$3,67 \pm 2,21$	$4,83 \pm 3,16$	$5,9 \pm 4,1$
Waxman	$1,74 \pm 0,67$	$2,45 \pm 3,02$	$2,05 \pm 3,11$	$1,9 \pm 3,2$

Fonte: Elaborada pelo autor.

Para a regra B-R1, o caminhante possui maior trajetória nas redes de pequeno-mundo, para todos os valores de memória avaliados, como mostram as tabelas 9 e 10. Em seguida, as maiores trajetórias são para as redes de Waxman e redes aleatórias. Para as redes de Barabasi-Albert, as trajetórias são curtas e o turista encontra rapidamente o

atrator ou, especialmente para valores elevados de memória, a trajetória é interrompida antes mesmo de o turista encontrar algum atrator.

Para a regra B-R2 (ver tabelas 9 e 10), as maiores trajetórias ainda são produzidas em redes de Watts-Strogatz. Essa regra produziu trajetórias maiores nas redes de Erdos-Renyi e trajetórias menores nas redes de Waxman se comparado à regra B-R1. As trajetórias nas redes do modelo BA continuam curtas.

Tabela 11 – Média e o desvio padrão dos transientes. Regra: B-R2.

Modelo	$t(\mu = 1)$	$t(\mu = 2)$	$t(\mu = 3)$	$t(\mu = 4)$
Barabasi-Albert	$0,02 \pm 0,1$	$0,04 \pm 0,07$	$0,01 \pm 0,07$	$0,02 \pm 0,06$
Erdos-Renyi	$0,42 \pm 0,56$	$5,92 \pm 10,8$	$6,4 \pm 10,79$	$6,8 \pm 10,5$
Watts-Strogatz	$2,27 \pm 1,59$	$4,45 \pm 7,73$	$4,51 \pm 7,95$	$6,1 \pm 11,7$
Waxman	$0,08 \pm 0,29$	$0,02 \pm 0,09$	$0,01 \pm 0,04$	$0,04 \pm 0,12$

Fonte: Elaborada pelo autor.

Tabela 12 – Média e o desvio padrão dos atratores. Regra: B-R2.

Modelo	$a(\mu = 1)$	$a(\mu = 2)$	$a(\mu = 3)$	$a(\mu = 4)$
Barabasi-Albert	$0,12 \pm 0,37$	$0,007 \pm 0,007$	$0,001 \pm 0,002$	$0,001 \pm 0,004$
Erdos-Renyi	$1,98 \pm 0,11$	$1,43 \pm 1,34$	$1,66 \pm 3,29$	$1,53 \pm 3,4$
Watts-Strogatz	$1,62 \pm 0,65$	$2,5 \pm 2,0$	$2,45 \pm 2,03$	$2,54 \pm 2,32$
Waxman	$1,91 \pm 0,38$	$0,08 \pm 0,21$	$0,008 \pm 0,024$	$0,004 \pm 0,011$

Fonte: Elaborada pelo autor.

Tabela 13 – Média e o desvio padrão dos transientes. Regra: B-R3.

Modelo	$t(\mu = 1)$	$t(\mu = 2)$	$t(\mu = 3)$	$t(\mu = 4)$
Barabasi-Albert	$0,50 \pm 0,86$	$17,18 \pm 15,32$	$48,42 \pm 21,29$	$22,93 \pm 18,03$
Erdos-Renyi	$0,38 \pm 0,65$	$19,10 \pm 12,03$	$24,54 \pm 19,84$	$14,65 \pm 12,25$
Watts-Strogatz	$0,61 \pm 1,09$	$10,74 \pm 9,63$	$50,61 \pm 34,23$	$23,84 \pm 26,80$
Waxman	$0,42 \pm 0,65$	$20,27 \pm 12,87$	$11,28 \pm 8,31$	$39,54 \pm 23,22$

Fonte: Elaborada pelo autor.

As regras B-R3 e B-R4 produzem trajetórias bem mais complexas, como mostram as Tabelas 13, 14, 15 e 16. Para turistas com memória $\mu = 1$, a caminhada é simples, são necessários poucos passos para encontrar um atrator e encerrar o movimento, bem como os atratores são menores, tendendo para o menor atrator possível $a = \mu + 1$. No entanto,

para valores de memória maiores do que um, o turista precisa percorrer um longo caminho até encontrar um atrator. O transiente e o atrator são métricas mais relevantes para as regras B-R3 e B-R4 do que o grau de bifurcação e de modo oposto, para as regras B-R1 e B-R2, o grau de bifurcação é uma métrica mais significativa do que a trajetória do turista.

Tabela 14 – Média e o desvio padrão dos atratores. Regra: B-R3.

Modelo	$a(\mu = 1)$	$a(\mu = 2)$	$a(\mu = 3)$	$a(\mu = 4)$
Barabasi-Albert	$2,00 \pm 0,00$	$19,66 \pm 2,15$	$10,46 \pm 1,39$	$19,01 \pm 11,79$
Erdos-Renyi	$2,00 \pm 0,00$	$43,12 \pm 7,13$	$5,91 \pm 2,79$	$38,59 \pm 7,41$
Watts-Strogatz	$2,00 \pm 0,12$	$3,06 \pm 0,25$	$5,18 \pm 0,97$	$34,64 \pm 32,55$
Waxman	$2,00 \pm 0,12$	$5,63 \pm 1,40$	$13,36 \pm 4,05$	$17,56 \pm 3,30$

Fonte: Elaborada pelo autor.

Para a regra B-R4, as redes aleatórias (ER), embora apresente baixo transiente, apresenta atratores bastante complexos, mostrando que o turista encontra certa dificuldade de encontrar atratores nestas redes. Devido às propriedades das redes livres de escala, o algoritmo usando a regra B-R4 consegue rapidamente encontrar um atrator, sendo os atratores menores comparado aos atratores produzidos pelo algoritmo nos outros modelos de rede. O mesmo não ocorre para a regra B-R3 nesta mesma rede.

Tabela 15 – Média e o desvio padrão dos transientes. Regra: B-R4.

Modelo	$t(\mu = 1)$	$t(\mu = 2)$	$t(\mu = 3)$	$t(\mu = 4)$
Barabasi-Albert	$1,90 \pm 1,68$	$5,75 \pm 4,00$	$2,92 \pm 2,59$	$10,91 \pm 5,44$
Erdos-Renyi	$0,45 \pm 0,72$	$10,63 \pm 7,88$	$11,11 \pm 8,48$	$12,84 \pm 14,13$
Watts-Strogatz	$0,82 \pm 1,72$	$8,38 \pm 7,39$	$7,97 \pm 6,76$	$20,11 \pm 15,32$
Waxman	$0,46 \pm 0,78$	$7,85 \pm 6,59$	$11,32 \pm 7,11$	$29,73 \pm 14,76$

Fonte: Elaborada pelo autor.

Tabela 16 – Média e o desvio padrão dos atratores. Regra: B-R4.

Modelo	$a(\mu = 1)$	$a(\mu = 2)$	$a(\mu = 3)$	$a(\mu = 4)$
Barabasi-Albert	$2,00 \pm 0,00$	$3,98 \pm 1,36$	$4,00 \pm 0,00$	$6,00 \pm 0,00$
Erdos-Renyi	$2,00 \pm 0,00$	$24,16 \pm 7,53$	$35,99 \pm 0,37$	$25,56 \pm 18,94$
Watts-Strogatz	$2,00 \pm 0,00$	$3,50 \pm 0,76$	$5,34 \pm 2,25$	$32,28 \pm 11,21$
Waxman	$2,00 \pm 0,00$	$44,88 \pm 2,10$	$10,83 \pm 0,89$	$6,00 \pm 0,00$

Fonte: Elaborada pelo autor.

6.3 O vetor de característica: Outras medidas para o método DTWB

Nesta seção será apresentado todas as medidas testadas para compor o vetor de característica. Ao decorrer do projeto diversas medidas foram testadas, algumas provaram-se boas medidas, outras ruins e com o processo de tentativa e erro foi construído o vetor de característica apresentado na Seção 4.1. Não apenas diversas medidas foram testadas, mas também diferentes combinações entre elas. O vetor apresentado na Seção 4.1 mostrou o melhor desempenho dentre todos os vetores testados, como mostra a Tabela 19. O *Dataset* 1, das redes sintéticas, foi utilizado como *dataset* de referência para testar as medidas. O valor de memória utilizado nos testes foi $\mu = 1$, o qual apresentou melhor desempenho, tanto do ponto de vista da acurácia quanto de custo computacional, dentre os valores individuais de memória.

A primeira tentativa de construir um vetor de característica para o método DTWB foi, a exemplo do DTW tradicional, a partir do histograma de alguma medida. Três vetores foram considerados: O vetor construído a partir do histograma de grau de bifurcação ($\phi(B)$), o vetor construído a partir do histograma de densidade de bifurcação ($\phi(\beta)$) e o vetor construído através do histograma conjunto de transiente e atrator ($\phi(\ell)$), assim como na caminhada do turista original (ver Equação 3.3). O primeiro possui 100 elementos, o segundo, 50 elementos e o terceiro possui 100 elementos. O desempenho das três medidas são apresentadas na Tabela 17 para as quatro regras de caminhada. O vetor $\phi(\beta)$ apresentou o melhor desempenho, seguido pelo vetor $\phi(B)$ e por fim, o vetor do DTW tradicional $\phi(\ell)$. No entanto, as três medidas tiveram um desempenho inferior aos métodos comparativos mostrados na Tabela 20.

Após testar o desempenho dos vetores extraídos das distribuições, foi testado inúmeras métricas e diversas combinações entre elas. Foram consideradas as seguintes métricas:

- Total de bifurcações na rede: soma dos graus de bifurcações de cada vértice
- Idade do turista: quantidade de iterações entre o início da trajetória e o fim. (Não confundir com o transiente. A idade considera o número de iterações a partir da bifurcação da trajetória, o transiente toma todas as iterações desde o nó de lançamento.)
- α : soma dos graus de bifurcação em uma trajetória sobre o comprimento desta trajetória.
- Grau de bifurcação (B) - Ver Capítulo 4
- Densidade de bifurcações (β) - Ver Capítulo 4
- Grau de não-bifurcação (!B): O oposto do grau de bifurcação.

- Transiente - Ver Capítulo 3
- Atrator - Ver Capítulo 3
- Quantidade de atratores comuns. Quantidade de trajetórias clones que caem sobre o mesmo atrator.

Das métricas acima, foram tomadas 24 medidas:

1. Total de bifurcações
2. Média de idade do turista
3. Média de α
4. Média de grau de bifurcação
5. Média de densidade de bifurcações
6. Média de grau de não-bifurcação
7. Média de transiente
8. Média de atrator
9. Quantidade atratores comuns
10. Quantidade de incidências no maior atrator comum.
11. Desvio padrão de idade do turista
12. Desvio padrão de α
13. Desvio padrão de grau de bifurcação
14. Desvio padrão de densidade de bifurcações
15. Desvio padrão de grau de não-bifurcação
16. Desvio padrão de transiente
17. Desvio de atrator
18. Entropia de idade do turista
19. Entropia de α
20. Entropia de grau de bifurcação
21. Entropia de densidade de bifurcações

22. Entropia de grau de não-bifurcação
23. Entropia de transiente
24. Entropia de atrator

Tabela 17 – Desempenho do vetor de característica gerado através dos histogramas. $\mu = 1$.
Redes do *Dataset 1*

Vetor de característica	B-R1	B-R2	B-R3	B-R4
$\phi(B)$	93,4%	92,1%	79,5%	88,0%
$\phi(\beta)$	95,7%	93,4%	82,3%	89,5%
$\phi(\ell)$	63,1%	66,6%	74,2%	81,4%

Fonte: Elaborada pelo autor.

Tabela 18 – Desempenho individual (%) das medidas utilizadas nos vetores de característica. $\mu = 1$. *Dataset*: Redes sintéticas.

Medida	B-R1	B-R2	B-R3	B-R4
Média da densidade de bifurcação	60,1%	60,9%	33,1%	46,5%
Média do grau de bifurcação	58,8%	55,4%	35,5%	43,1%
Média do transiente	58,4%	54,6%	45,2%	53,4%
Média do atrator	57,7%	56,9%	40,0%	49,9%
Total de bifurcações	56,1%	52,4%	33,1%	35,6%
Desvio padrão do grau de bifurcação	55,5%	55,1%	34,6%	45,1%
Desvio padrão do transiente	55,3%	53,7%	44,9%	45,6%
Média de α	55,1%	48,6%	36,7%	42,4%
Desvio padrão de α	54,6%	54,1%	37,2%	38,5%
Entropia do grau de bifurcação	54,3%	54,4%	36,8%	38,0%
Entropia da densidade de bifurcação	53,4%	52,7%	35,9%	38,0%
Desvio padrão do transiente	53,1%	52,1%	37,3%	46,7%
Desvio padrão do atrator	53,1%	52,2%	37,6%	46,1%
Entropia do transiente	51,5%	51,2%	36,1%	43,8%
Entropia do atrator	50,2%	50,0%	36,4%	44,0%
Média do grau de não-bifurcação	49,1%	48,9%	32,5%	33,9%
Entropia de α	47,8%	43,9%	31,3%	33,9%
Média de idade do turista	43,5%	42,1%	30,7%	33,2%
Desvio padrão da idade do turista	41,8%	41,3%	31,9%	31,7%
Entropia da idade do turista	41,4%	41,9%	31,7%	31,1%
Desvio padrão do grau de não-bifurcação	41,4%	40,1%	30,0%	31,2%
Entropia do grau de não-bifurcação	40,1%	36,6%	27,1%	32,2%
Quantidade de atratores comuns	30,0%	35,6%	25,2%	28,3%
Incidência do menor atrator comum	29,9%	29,6%	25,3%	28,6%

Fonte: Elaborada pelo autor.

Tabela 19 – Desempenho (%) das combinações de medidas no vetor de característica.
 $\mu = 1$.

Medidas	B-R1	B-R2	B-R3	B-R4
[4, 5, 7, 8, 13, 14, 16, 17, 20, 21, 23, 24]	97,0%	95,9%	93,1%	96,5
[4, 5, 7, 8, 13, 14, 16, 17, 20, 21, 23]	96,5%	95,4%	95,5%	93,1%
[4, 5, 7, 8, 13, 14, 16, 17, 20, 21]	95,5%	94,6%	95,2%	93,4%
[4, 5, 7, 8, 13, 14, 16, 17, 20]	90,1%	88,9%	90,0%	89,9%
[4, 5, 7, 8, 13, 14, 16, 17]	86,0%	82,4%	83,1%	85,6%
[4, 5, 7, 8, 13, 14, 16]	78,4%	75,1%	74,6%	75,1%
[4, 5, 7, 8, 13, 14]	65,8%	63,7%	64,9%	65,6%
[4, 5]	60,9%	58,6%	58,7%	59,4%
[4, 13, 21]	58,6%	55,1%	57,2%	58,5%
[4, 13]	58,3%	55,4%	56,8%	58,0%
[7, 8, 16, 17, 23, 24]	58,3%	54,7%	55,9%	58,0%
[7,16,23]	57,5%	54,1%	54,3%	56,7%
[8,17,24]	57,1%	54,2%	54,6%	56,9%
[4, 5, 7, 8, 13]	54,5%	51,2%	51,1%	53,8%
[1, 4, 5, 7, 8, 13]	54,2%	51,0%	51,0%	54,0%
[7, 8]	54,1%	50,9%	50,5%	53,9%
[1, 2, 4, 5, 7, 8, 13]	53,8%	50,9%	51,3%	53,9%
[1, 2, 4, 5, 7, 8, 13, 14]	53,5%	50,1%	50,0%	53,2%
[1, 2, 4, 5, 7, 8, 13, 14, 15]	51,8%	48,3%	47,9%	51,7%
[1, 2, 4, 5, 7, 8, 13, 14, 15, 16, 17]	51,4%	48,9%	48,7%	51,1%
[7, 8, 11]	51,4%	48,1%	48,0%	51,2%
[1, 2, 4, 5, 7, 8, 13, 14, 15, 16, 17, 20, 21]	50,1%	46,6%	47,1%	48,2%
[1, 2, 4, 5, 7, 8, 13, 14, 15, 16, 17, 20]	50,0%	45,6%	45,2%	48,3%
[1, 2]	48,9%	45,6%	45,3%	48,6%
[1, 2, 3, 4]	48,4%	44,9%	45,3%	48,3%
[1, 9]	48,4%	44,2%	44,7%	47,9%
[1, 7]	48,0%	44,1%	44,3%	47,6%
[1, 4, 13]	47,5%	44,1%	44,1%	47,3%
[1, 2, 3, 4, 5]	47,0%	43,5%	43,2%	46,1%
[3, 9, 17, 18, 19]	45,0%	43,4%	43,1%	44,4%
[3, 12, 19]	41,3%	33,7%	34,5%	40,6%
[3, 7, 8]	41,3%	33,3%	33,8%	40,2%
[3, 4, 5, 12, 16]	41,0%	33,1%	33,0%	40,3%
[10, 16, 17, 18]	41,0%	32,6%	33,0%	40,2%
[7, 8, 17, 20, 21, 22]	40,4%	32,6%	32,5%	40,0%
[6, 9, 10, 11, 12, 19]	40,3%	32,0%	32,5%	40,2%
[1, ..., 24]	40,0%	32,6%	32,8%	40,2%
[17, 24]	39,8%	30,6%	27,1%	39,9%
[16, 23]	39,0%	30,1%	27,9%	39,6%
[21, 23]	38,5%	30,2%	27,9%	38,7%
[13, 19, 23, 24]	37,6%	29,7%	29,3%	37,7%
[20, 21, 22]	37,6%	29,0%	29,1%	37,3%

Fonte: Elaborada pelo autor.

A Tabela 18 mostra o desempenho individual de cada medida. A medida 5 (Média de β) apresentou melhor resultado individual, acumulando sozinha 60% de acurácia na classificação de redes sintéticas para as regras B-R1 e B-R2. Em seguida está a medida 4 (média do grau de bifurcação) e em seguida a medida 7 (média do transiente). Em último lugar está a medida 10, quantidade de incidências do maior atrator comum. A Tabela 19 mostra o desempenho das diferentes combinações de medidas para as quadro regras de caminhada e memória $\mu = 1$.

6.4 Reconhecimento de padrões

Nesta seção será apresentado o desempenho do método da caminhada determinística do turista com bifurcações na classificação de redes sintéticas e de redes reais comparado com outros métodos mostrados na Tabela 20. Foram usados três métodos para comparação (ver Seção 5.3), o LLNA (*Binary pattern*), medidas estruturais e a caminhada determinística do turista (DTW) usando a combinação de quatro valores de memória: $\vec{\mu} = [1, 2, 3, 4]$ para as regras N-R3, N-R4 e para a combinação de ambas.

Tabela 20 – Valores de acurácia dos métodos utilizados como comparação ao método DTWB. o Método DTW usado nos experimentos utiliza quatro valores de memória combinados ($\vec{\mu} = [1, 2, 3, 4]$) para cada regra de caminhada.

<i>Dataset</i>	LLNA-BP	Medidas estruturais	DTW tradicional		
			N-R3	N-R4	N-R3 \cup N-R4
<i>dataset 1</i>	100,0%	100,0%	95,3%	97,0%	98,5%
<i>dataset 2</i> ($\rho = 10\%$)	100,0%	100,0%	95,0%	96,5%	96,7%
<i>dataset 2</i> ($\rho = 20\%$)	99,9%	100,0%	94,7%	95,8%	96,5%
<i>dataset 2</i> ($\rho = 30\%$)	99,9%	100,0%	94,1%	95,2%	96,3%
<i>dataset 4</i> (Actinobactéria)	95,1%	93,1%	90,0%	94,0%	94,9%
<i>dataset 4</i> (Animais)	84,8%	83,7%	78,1%	75,4%	80,0%
<i>dataset 4</i> (Firmicutes)	98,3%	95,6%	90,3%	90,7%	93,3%
<i>dataset 4</i> (Fungi)	76,7%	54,9%	62,7%	68,1%	68,9%
<i>dataset 4</i> (Plantas)	74,8%	54,1%	57,7%	57,7%	59,8%
<i>dataset 4</i> (Protistas)	87,0%	45,1%	57,8%	57,8%	61,4%
<i>dataset 4</i> (Reinos)	97,4%	96,6%	83,6%	87,2%	89,6%
<i>dataset 5</i>	93,4%	88,0%	68,6%	71,7%	74,5%

Fonte: RIBAS *et al.* (31)

A Tabela 20 mostra que tanto o método LLNA quanto as medidas estruturais possuem robustez resultando em um bom desempenho mesmo para redes com ruído, essa tabela também mostra que a acurácia do método DTW tradicional decresce conforme é adicionado ruído na rede. O método DTW tradicional mostrou-se inferior ao método LLNA e, em alguns casos, ao método das medidas estruturais.

Tabela 21 – Valores de acurácia do método DTWB para valores únicos de memória e regra B-R1.

<i>Dataset</i>	$\mu = 1$	$\mu = 2$	$\mu = 3$	$\mu = 4$	$\mu = 5$	$\mu = 6$
<i>dataset 1</i>	97,0%	97,1%	97,2%	96,8%	97,0%	97,1%
<i>dataset 2</i> ($\rho = 10\%$)	96,3%	96,2%	96,1%	96,3%	96,2%	96,1%
<i>dataset 2</i> ($\rho = 20\%$)	95,7%	95,7%	95,3%	95,3%	95,6%	95,4%
<i>dataset 2</i> ($\rho = 30\%$)	94,9%	95,0%	94,8%	94,3%	94,9%	94,4%
<i>dataset 4</i> (Actinobactéria)	99,4%	99,4%	99,5%	99,5%	99,4%	99,4%
<i>dataset 4</i> (Animais)	94,6%	94,7%	94,6%	94,7%	94,6%	94,6%
<i>dataset 4</i> (Firmicutes)	97,3%	97,5%	97,5%	97,5%	97,8%	97,5%
<i>dataset 4</i> (Fungi)	78,3%	78,3%	78,3%	78,3%	78,5%	78,3%
<i>dataset 4</i> (Plantas)	81,4%	81,8%	81,4%	81,4%	81,4%	81,5%
<i>dataset 4</i> (Protistas)	85,0%	85,0%	85,1%	85,1%	85,0%	85,0%
<i>dataset 4</i> (Reinos)	94,3%	94,3%	94,3%	94,4%	94,5%	94,3%
<i>dataset 5</i>	81,0%	81,0%	81,0%	82,0%	82,0%	82,0%

Fonte: Elaborada pelo autor.

Tabela 22 – Valores de acurácia do método DTWB para valores únicos de memória e regra B-R2.

<i>Dataset</i>	$\mu = 1$	$\mu = 2$	$\mu = 3$	$\mu = 4$	$\mu = 5$	$\mu = 6$
<i>dataset 1</i>	95,9%	95,5%	95,6%	95,8%	95,8%	95,6%
<i>dataset 2</i> ($\rho = 10\%$)	95,8%	95,7%	95,8%	95,8%	95,9%	96,0%
<i>dataset 2</i> ($\rho = 20\%$)	95,2%	95,6%	95,4%	95,5%	95,6%	95,3%
<i>dataset 2</i> ($\rho = 30\%$)	95,3%	94,9%	95,0%	94,9%	94,7%	95,2%
<i>dataset 4</i> (Actinobactéria)	98,9%	98,9%	98,9%	98,9%	98,9%	98,9%
<i>dataset 4</i> (Animais)	98,2%	98,7%	98,5%	98,5%	98,5%	98,5%
<i>dataset 4</i> (Firmicutes)	93,4%	93,4%	93,4%	93,4%	93,3%	93,4%
<i>dataset 4</i> (Fungi)	81,6%	81,7%	81,6%	81,6%	81,6%	81,6%
<i>dataset 4</i> (Plantas)	92,5%	92,6%	92,6%	92,6%	92,6%	92,6%
<i>dataset 4</i> (Protistas)	65,0%	70,0%	70,1%	70,1%	70,0%	70,0%
<i>dataset 4</i> (Reinos)	98,1%	98,3%	98,1%	98,1%	98,1%	98,1%
<i>dataset 5</i>	78,0%	78,0%	78,0%	78,0%	78,0%	78,1%

Fonte: Elaborada pelo autor.

As Tabelas 21 e 22 mostram o desempenho do método DTWB para valores únicos de memória para as regras B-R1 e B-R2, respectivamente. Para essas regras, o método DTWB apresentou pouca variação de desempenho em relação ao valor de memória. A acurácia para redes sintéticas, com ruído e sem ruído, foi levemente inferior ao desempenho do LLNA e das medidas estruturais para ambas as regras e com desempenho equivalente à melhor versão do método DTW tradicional ($N-R3 \cup N-R4$). Desse modo, o método proposto, usando somente um valor de memória por execução, conseguiu se igualar a melhor versão de seu predecessor, o qual usa vários tamanhos de memória combinados. A

regra B-R2, embora tenha uma acurácia menor do que a regra B-R1 para redes sintéticas, apresentou uma maior resistência a ruídos, variando pouco conforme ruído é adicionado às redes. A Figura 17 mostra a classificação das redes sintéticas com ruído e sem ruído para as regras B-R1 e B-R2 utilizando o método de classificação PCA. A Figura 18 mostra a matrizes confusão para as redes sintéticas.

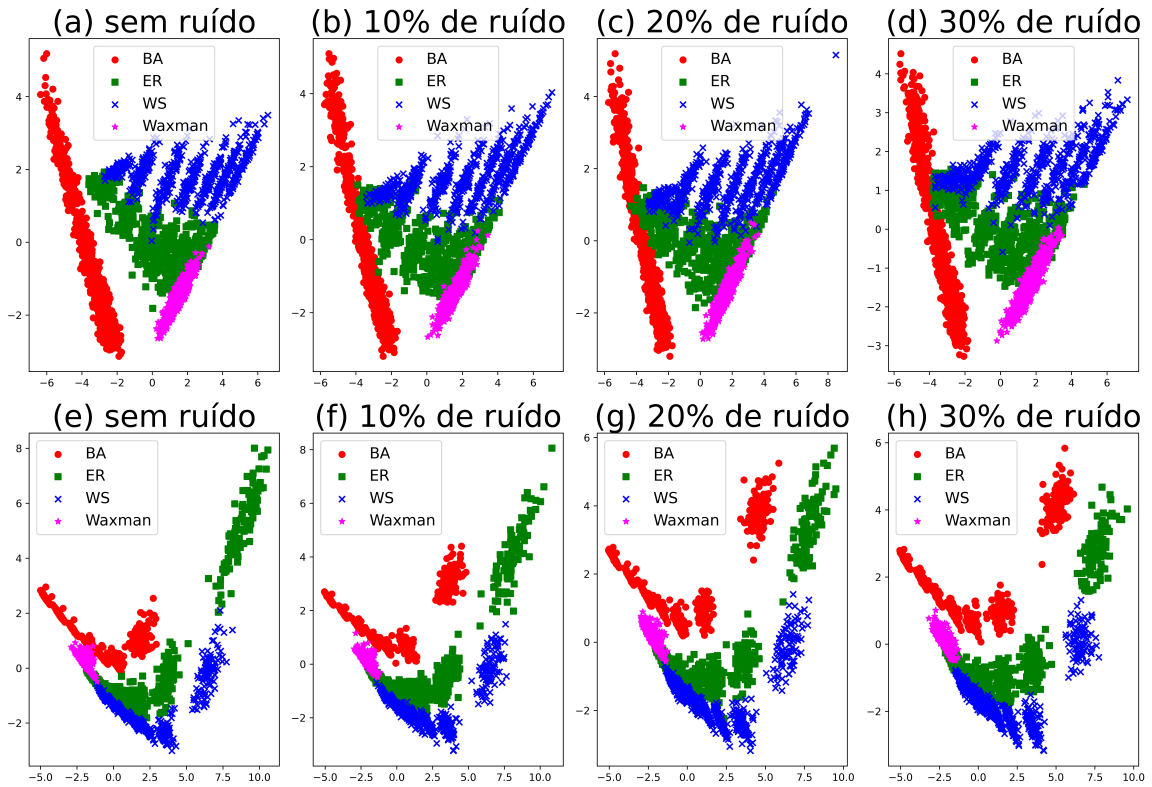


Figura 17 – Classificação através do método PCA usando os vetores de característica produzidos pelo método DTWB para as regras B-R1 (linha de cima) e B-R2 (linha de baixo). Os eixos são as componentes principais.

Fonte: Elaborada pelo autor.

Para a regra B-R1 o método superou o LLNA em quatro *datasets* reais: Actinobactéria, Animais, *Fungi* e Plantas, e obteve desempenho semelhante em um: Firmicutes-Bacillus, perdendo nos *datasets* Reinos, Protistas e redes sociais (*dataset* 5). Essa regra só obteve desempenho inferior ao das medidas estruturais para as redes sociais e o *dataset* Reinos. Já a regra B-R2 superou o LLNA em quatro *datasets* reais, teve desempenho semelhante em um (Reinos) e perdeu em três (Firmicutes-Bacillus, Protistas e redes sociais). Para as redes reais em alguns casos a regra B-R1 apresentou desempenho superior, em outros a regra B-R2 foi superior. As duas regras são antagônicas, sendo que a primeira favorece redes homogêneas e a segunda favorece redes heterogêneas. Portanto, isso nos permite escolher qual regra utilizar caso a caso, escolhendo aquela que provém o melhor desempenho para o *dataset* em questão.

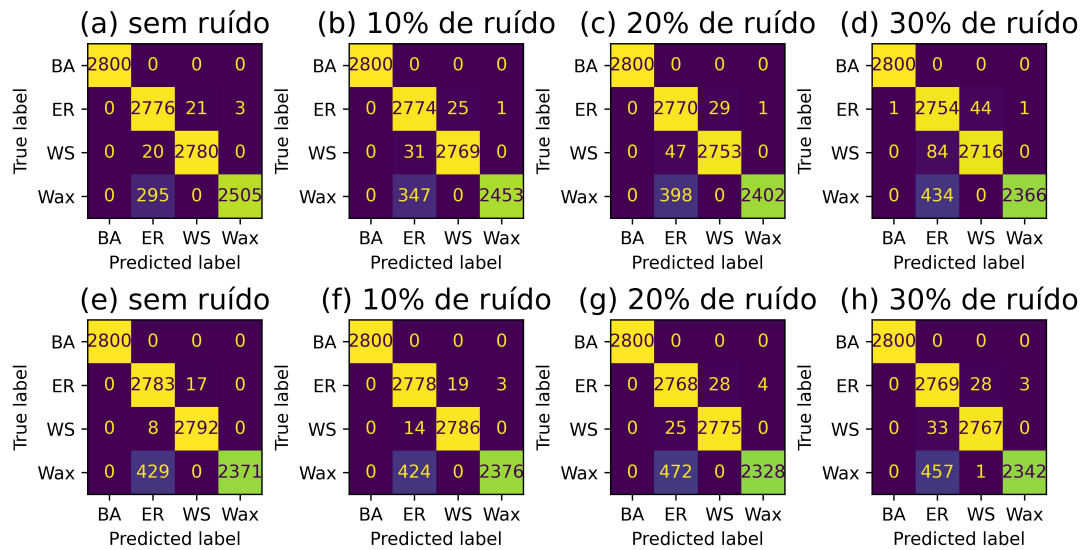


Figura 18 – Matrizes de confusão da classificação das redes do *dataset 1* e do *dataset 2* para as regras B-R1 (linha de cima) e B-R2 (linha de baixo).

Fonte: Elaborada pelo autor.

Tabela 23 – Valores de acurácia do método DTWB para valores únicos de memória e regra B-R3.

<i>Dataset</i>	$\mu = 1$	$\mu = 2$	$\mu = 3$	$\mu = 4$	$\mu = 5$	$\mu = 6$
<i>dataset 1</i>	93,1%	92,2%	92,2%	91,7%	91,0%	91,3%
<i>dataset 2</i> ($\rho = 10\%$)	92,6%	92,1%	92,3%	91,6%	91,7%	91,2%
<i>dataset 2</i> ($\rho = 20\%$)	90,6%	90,4%	90,0%	90,1%	90,3%	90,1%
<i>dataset 2</i> ($\rho = 30\%$)	89,1%	89,0%	88,9%	88,3%	88,1%	88,0%
<i>dataset 4</i> (Actinobactéria)	83,6%	57,3%	70,3%	72,3%	71,4%	70,6%
<i>dataset 4</i> (Animais)	82,1%	66,0%	60,7%	50,0%	44,9%	44,6%
<i>dataset 4</i> (Firmicutes)	90,5%	78,5%	76,3%	64,0%	59,8%	59,5%
<i>dataset 4</i> (Fungi)	53,3%	33,3%	40,0%	55,0%	54,5%	55,2%
<i>dataset 4</i> (Plantas)	48,1%	66,6%	51,8%	55,5%	55,4%	56,3%
<i>dataset 4</i> (Protistas)	80,0%	45,0%	60,0%	35,0%	40,2%	45,3%
<i>dataset 4</i> (Reinos)	89,3%	86,2%	75,0%	75,6%	75,0%	75,3%
<i>dataset 5</i>	66,0%	61,0%	66,0%	58,0%	59,0%	54,6%

Fonte: Elaborada pelo autor.

As Tabelas 23 e 24 mostram, respectivamente, os resultados para as regras B-R3 e B-R4. Os resultados para as duas regras foram inferiores aos resultados para as regras B-R1 e B-R2. A regra B-R3 obteve baixo desempenho, incluindo para redes sintéticas, e tal regra demonstrou uma alta variação de desempenho com relação ao valor de memória sendo esse, portanto, um parâmetro sensível do modelo para essa regra. A regra B-R4 obteve

um desempenho melhor, com pouca variação de acurácia conforme o valor de memória e para alguns *datasets* atingiu desempenho semelhante às regras B-R1 e B-R2 como, por exemplo, para os *datasets* Plantas (memória $\mu = 1$), Reinos e Protistas, onde superou as regras B-R1 e B-R2 e o método LLNA.

Tabela 24 – Valores de acurácia do método DTWB para valores únicos de memória e regra B-R4.

<i>Dataset</i>	$\mu = 1$	$\mu = 2$	$\mu = 3$	$\mu = 4$	$\mu = 5$	$\mu = 6$
<i>dataset 1</i>	96,1%	95,6%	95,7%	95,2%	95,1%	94,7%
<i>dataset 2</i> ($\rho = 10\%$)	94,9%	94,1%	94,2%	93,6%	93,8%	93,2%
<i>dataset 2</i> ($\rho = 20\%$)	94,0%	93,4%	93,4%	93,0%	93,1%	92,5%
<i>dataset 2</i> ($\rho = 30\%$)	93,1%	92,3%	92,0%	91,8%	91,7%	91,4%
<i>dataset 4</i> (Actinobactéria)	96,4%	96,9%	96,9%	95,9%	95,9%	95,8%
<i>dataset 4</i> (Animais)	92,8%	91,0%	91,5%	91,0%	90,9%	90,5%
<i>dataset 4</i> (Firmicutes)	86,9%	84,7%	86,7%	85,2%	85,3%	85,0%
<i>dataset 4</i> (Fungi)	78,3%	76,6%	75,0%	75,2%	74,8%	74,6%
<i>dataset 4</i> (Plantas)	92,5%	55,5%	55,0%	54,6%	55,5%	55,6%
<i>dataset 4</i> (Protistas)	90,0%	90,0%	70,0%	69,1%	68,4%	65,1%
<i>dataset 4</i> (Reinos)	95,6%	95,6%	95,1%	95,2%	94,9%	95,0%
<i>dataset 5</i>	57,0%	58,0%	62,0%	65,0%	63,0%	64,0%

Fonte: Elaborada pelo autor.

Tabela 25 – Valores de acurácia do método DTWB combinando valores de memória para a regra B-R1.

<i>Dataset</i>	[1, 2]	[1, 2, 3]	[1, 2, 3, 4]	[1, 2, 3, 4, 5]
<i>dataset 1</i>	98,2%	99,3%	99,6%	99,8%
<i>dataset 2</i> ($\rho = 10\%$)	98,1%	99,0%	99,5%	99,7%
<i>dataset 2</i> ($\rho = 20\%$)	97,7%	98,6%	99,1%	99,5%
<i>dataset 2</i> ($\rho = 30\%$)	97,3%	98,1%	98,7%	99,1%
<i>dataset 4</i> (Actinobactéria)	99,4%	99,4%	99,4%	99,4%
<i>dataset 4</i> (Animais)	94,6%	94,6%	94,6%	96,4%
<i>dataset 4</i> (Firmicutes)	94,2%	94,2%	94,2%	94,2%
<i>dataset 4</i> (Fungi)	78,3%	78,3%	78,3%	78,3%
<i>dataset 4</i> (Plantas)	81,4%	77,7%	77,7%	81,4%
<i>dataset 4</i> (Protistas)	85,0%	85,0%	85,0%	85,0%
<i>dataset 4</i> (Reinos)	95,0%	95,2%	95,0%	95,0%
<i>dataset 5</i>	83,0%	83,0%	83,1%	83,0%

Fonte: Elaborada pelo autor.

A união de diferentes valores de memória para as regras B-R1 e B-R2 usando o vetor de característica definido pela equação 3.4 não apresentou um ganho de desempenho para redes reais, como mostram as tabelas 25 e 26. No entanto, houve ganho de desempenho

para redes sintéticas, especialmente àquelas com ruído. A pouca variação de desempenho, tanto para valores singulares de memória quanto para combinações destes, mostra que o método DTWB com as regras B-R1 e B-R2 não possui grande dependência do valor de memória quanto ao método DTW original e ao método DTWB com as regras B-R3 e B-R4.

Tabela 26 – Valores de acurácia do método DTWB combinando valores de memória para a regra B-R2.

<i>Dataset</i>	[1, 2]	[1, 2, 3]	[1, 2, 3, 4]	[1, 2, 3, 4, 5]
<i>dataset 1</i>	98,2%	99,3%	99,6%	99,8%
<i>dataset 2</i> ($\rho = 10\%$)	96,5%	96,8%	97,2%	97,4%
<i>dataset 2</i> ($\rho = 20\%$)	96,8%	97,7%	98,2%	98,5%
<i>dataset 2</i> ($\rho = 30\%$)	96,6%	97,2%	97,9%	98,3%
<i>dataset 4</i> (Actinobactéria)	98,9%	98,9%	98,9%	98,9%
<i>dataset 4</i> (Animais)	98,2%	98,2%	98,2%	98,2%
<i>dataset 4</i> (Firmicutes)	94,2%	94,2%	94,2%	94,2%
<i>dataset 4</i> (Fungi)	81,6%	81,6%	81,6%	81,7%
<i>dataset 4</i> (Plantas)	92,5%	92,5%	92,6%	92,6%
<i>dataset 4</i> (Protistas)	65,0%	70,0%	70,0%	70,0%
<i>dataset 4</i> (Reinos)	98,1%	97,5%	97,5%	97,5%
<i>dataset 5</i>	81,0%	81,0%	81,0%	81,0%

Fonte: Elaborada pelo autor.

Tabela 27 – Valores de acurácia do método DTWB combinando valores de memória para a regra B-R3.

<i>Dataset</i>	[1, 2]	[1, 2, 3]	[1, 2, 3, 4]	[1, 2, 3, 4, 5]
<i>dataset 1</i>	94,2%	94,3%	94,8%	95,9%
<i>dataset 2</i> ($\rho = 10\%$)	93,1%	93,2%	93,2%	93,6%
<i>dataset 2</i> ($\rho = 20\%$)	91,4%	91,7%	92,3%	93,5%
<i>dataset 2</i> ($\rho = 30\%$)	89,4%	90,2%	90,7%	91,2%
<i>dataset 4</i> (Actinobactéria)	86,4%	87,4%	87,9%	88,4%
<i>dataset 4</i> (Animais)	89,2%	87,5%	85,7%	85,4%
<i>dataset 4</i> (Firmicutes)	92,9%	92,5%	94,6%	94,9%
<i>dataset 4</i> (Fungi)	51,6%	55,0%	56,6%	57,9%
<i>dataset 4</i> (Plantas)	66,6%	62,9%	40,0%	40,2%
<i>dataset 4</i> (Protistas)	55,0%	75,0%	90,0%	92,0%
<i>dataset 4</i> (Reinos)	93,7%	91,8%	92,5%	93,4%
<i>dataset 5</i>	61,0%	63,0%	64,1%	83,0%

Fonte: Elaborada pelo autor.

Já para as regras B-R3 e B-R4 houve melhora no desempenho do método devido à concatenação de vetores de característica para diferentes tamanhos de memória somente

para as redes sintéticas. Para as redes reais, em alguns casos, houve queda de desempenho. A regra B-R4 foi novamente superior à regra B-R3, mas ainda é inferior às regras B-R1 e B-R2 na maioria dos casos, demonstrando novamente que a memória é um parâmetro significativo para tais regras. As Tabelas 27 e 28 mostram o desempenho das regras B-R3 e B-R4 com a combinação de tamanhos de memória.

Tabela 28 – Valores de acurácia do método DTWB combinando valores de memória para a regra B-R4.

<i>Dataset</i>	[1, 2]	[1, 2, 3]	[1, 2, 3, 4]	[1, 2, 3, 4, 5]
<i>dataset 1</i>	96,8%	97,3%	97,9%	98,7%
<i>dataset 2</i> ($\rho = 10\%$)	96,2%	96,7%	97,2%	97,8%
<i>dataset 2</i> ($\rho = 20\%$)	94,6%	94,9%	95,6%	96,5%
<i>dataset 2</i> ($\rho = 30\%$)	94,5%	94,8%	95,9%	96,2%
<i>dataset 4</i> (Actinobactéria)	96,9%	97,4%	98,4%	98,8%
<i>dataset 4</i> (Animais)	94,6%	94,9%	96,8%	97,2%
<i>dataset 4</i> (Firmicutes)	91,5%	91,7%	91,9%	92,0%
<i>dataset 4</i> (Fungi)	81,6%	86,6%	86,6%	87,1%
<i>dataset 4</i> (Plantas)	66,6%	66,5%	67,6%	67,6%
<i>dataset 4</i> (Protistas)	75,0%	75,0%	70,0%	72,0%
<i>dataset 4</i> (Reinos)	95,0%	95,1%	96,5%	96,8%
<i>dataset 5</i>	62,0%	61,0%	59,0%	55,0%

Fonte: Elaborada pelo autor.

Tabela 29 – Valores de acurácia do método DTWB combinando as regras de caminhada. Em cada caso também há combinação de memória: $\vec{\mu} = [1, 2, 3, 4]$. Comb1 = B-R1 \cup B-R2, Comb2 = B-R3 \cup B-R4 e Comb3 = B-R1 \cup B-R2 \cup B-R3 \cup B-R4.

<i>Dataset</i>	Comb1	Comb2	Comb3
<i>dataset 1</i>	99,9%	97,2%	98,9%
<i>dataset 2</i> ($\rho = 10\%$)	99,9%	97,0%	98,1%
<i>dataset 2</i> ($\rho = 20\%$)	99,9%	96,8%	97,9%
<i>dataset 2</i> ($\rho = 30\%$)	99,8%	96,4%	97,1%
<i>dataset 4</i> (Actinobactéria)	99,4%	98,9%	99,2%
<i>dataset 4</i> (Animais)	98,4%	87,5%	94,4%
<i>dataset 4</i> (Firmicutes)	97,3%	96,3%	97,2%
<i>dataset 4</i> (Fungi)	86,6%	72,0%	81,2%
<i>dataset 4</i> (Plantas)	92,5%	68,8%	74,1%
<i>dataset 4</i> (Protistas)	85,9%	75,0%	79,1%
<i>dataset 4</i> (Reinos)	98,7%	95,6%	97,9%
<i>dataset 5</i>	85,9%	64,0%	69,0%

Fonte: Elaborada pelo autor.

Quando combinamos as duas regras como definido pela equação 3.5, o desempenho do método para as regras B-R1 e B-R2 aumenta, como mostra a tabela 29. Já a união das regras B-R3 e B-R4 atingiu desempenho inferior em todos os casos e quando combinamos as quatro regras, há uma queda de desempenho com relação à união $B - R1 \cup B - R2$, devido à influência da baixa performance das regras B-R3 e B-R4.

A união das regras B-R1 e B-R2 gerou um desempenho equivalente ao desempenho do LLNA e das medidas estruturais para redes sintéticas do *dataset* 1 e redes sintéticas com ruído do *dataset* 2, atingindo acurácia superior a 99%. Em redes reais, o método proposto, com as regras B-R1 e B-R2, obteve melhor desempenho em relação às medidas estruturais em quase todos os experimentos, a exceção é o *dataset* das redes sociais. O método DTWB unindo as duas regras foi melhor que o LLNA exceto para o *dataset* Firmicutes-Bacillus, onde a diferença é baixa, assim como no *dataset* das redes sociais.

As regras B-R3 e B-R4 obtiveram desempenho inferior às regras B-R1 e B-R2 na maioria dos casos, tanto para valores singulares de memória quanto para combinações de memórias. O mesmo vale para a combinação $B - R3 \cup B - R4$, o qual atingiu menor acurácia do que o par $B - R1 \cup B - R2$. Dessa forma, é seguro dizer, pelos resultados obtidos e apresentados aqui, que as regras B-R1 e B-R2 são mais adequadas ao método DTWB, sendo o maior desempenho atingindo com a união de memórias e regras. O método DTWB com as regras B-R1 e B-R2 são pouco sensíveis ao tamanho de memória, como mostram as Tabelas 21 e 22 e pouco dependente das trajetórias, como mostram os resultados da Seção 6.2.2, sendo os valores de grau de bifurcação e densidade de bifurcação preponderantes para esse método, como mostram os resultados das Seções 6.2.1 e 6.3.

6.5 Métricas de Pequeno-mundo baseadas na caminhada do turista

Para construir uma métrica para redes de pequeno-mundo foi utilizado o método discutido na Seção 3.3 e representado pela Equação 3.7. Executando o método para 500 redes de Watts-Strogatz, para probabilidade de religação variando entre 10^{-4} e 10^0 , os vetores de característica, definidos pela Equação 3.6, formam um padrão, como mostra a Figura 19, indo de redes quase regulares ($p \approx 10^{-4}$) até redes aleatórias (10^0). Os gráficos de dispersão da Figura 19 foram produzidos com auxílio do método PCA para a redução de dimensionalidade dos vetores e os eixos x e y representam as duas principais componentes do PCA.

Note que há uma transição entre os dois extremos (regular e aleatório), incluindo uma mudança na direção de dispersão dos vetores. Essa mudança de direção ocorre entre $p \approx 10^{-2}$ e $p \approx 2 \cdot 10^{-1}$, faixa de probabilidade onde estão as redes de pequeno-mundo, conforme as outras métricas. Usando a métrica ω a fim de comparação, tal métrica tende à zero em $p \approx 0.11$ para grau médio igual a 6 e tende a zero em $p \approx 0.09$ para $\langle k \rangle = 10$. Os dois valores estão no intervalo citado.

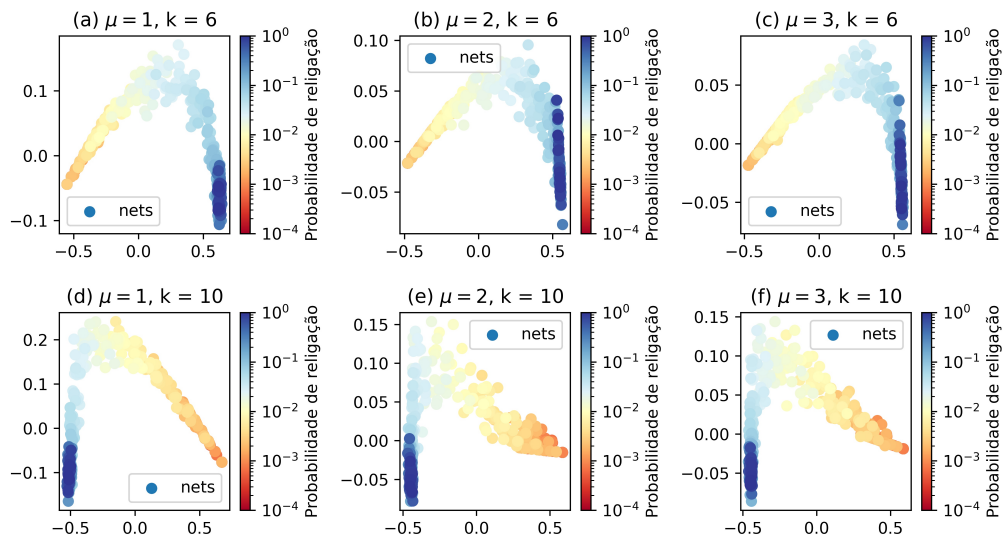


Figura 19 – *Scatter plot* dos vetores de característica (Eq. 3.6) para redes de Watts-Strogatz com probabilidade variando entre 10^{-4} (vermelho) e 10^0 (azul).

Fonte: Elaborada pelo autor.

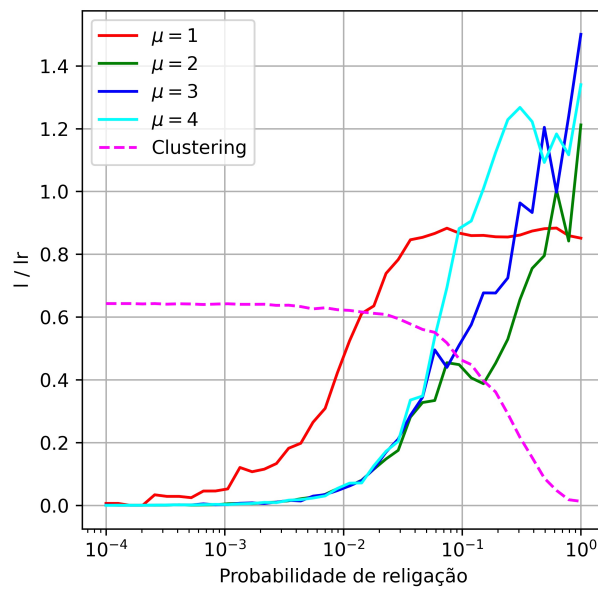


Figura 20 – Valores de comprimento da trajetória do turista (l), normalizado pelo comprimento da trajetória do turista para uma rede aleatória equivalente (l_r). Foram usados quatro valores de memória: $\mu = 1, 2, 3, 4$. A linha tracejada em magenta mostra a evolução do coeficiente de aglomeração no mesmo intervalo de probabilidade de religação. Redes com $N = 500$ e $\langle k \rangle = 10$.

Fonte: Elaborada pelo autor.

Usando esse método como uma métrica de pequeno-mundo através do coeficiente χ (Eq. 3.7) esperamos que $\chi \rightarrow 0$ para redes aleatórias e regulares e tenha valores positivos (um pico) para redes de pequeno-mundo, uma vez que o comprimento das trajetórias do turista tende ao valor máximo para redes aleatórias e o mesmo anula-se para redes regulares, enquanto o coeficiente de aglomeração anula-se para redes aleatórias, como mostra a Figura 20, tendendo, portanto, ao pico quando as duas medidas (comprimento da trajetória e coeficiente de aglomeração) estão suficientemente altas. A Figura 20 também mostra que o comprimento das trajetórias do turista aumenta no referido intervalo ($[10^{-2}, 2.10^{-1}]$).

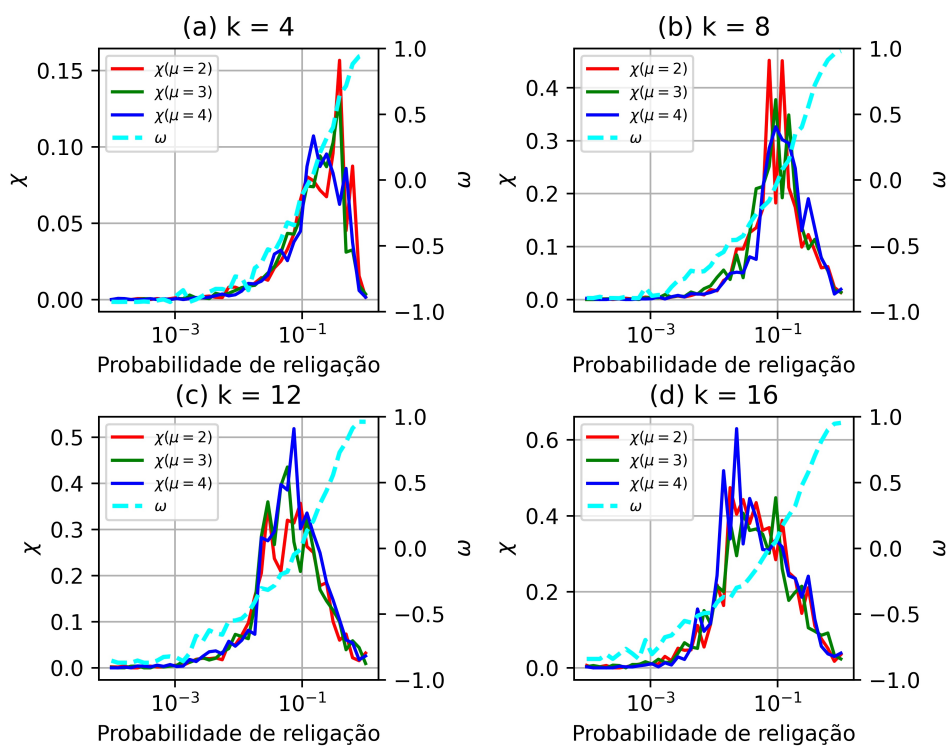


Figura 21 – Valores de χ (eixo esquerdo) e ω (eixo direito) para redes de Watts-Strogatz para $p \in [10^{-4}, 10^0]$. As linhas sólidas em vermelho, verde e azul são os valores de χ para memória $\mu = 2$, $\mu = 3$ e $\mu = 4$, respectivamente. A linha tracejada em ciano são os valores do índice ω . Todas as redes possuem 500 nós.

Fonte: Elaborada pelo autor.

A Figura 21 mostra o valor do coeficiente χ comparado ao valor do índice ω para redes de Watts-Strogatz no intervalo $[10^{-4}, 10^0]$ de probabilidade de religação. Note que o pico das curvas do coeficiente χ ocorrem dentro intervalo $[10^{-2}, 2.10^{-1}]$ e próximo ao ponto onde o índice ω é igual a zero e também dentro do intervalo sugerido na Seção 2.2.3 com $-0.5 \leq \omega \leq 0.5$. Para $k = 4$, $k = 8$, $k = 12$ e $k = 16$ o valor de ω se anula em aproximadamente $p = 0,14$, $p = 0,10$, $p = 0,09$ e $p = 0,08$, respectivamente. As redes em torno destas probabilidades, considerando os respectivos graus médios, exibem

propriedades de rede de pequeno-mundo.

A Tabela 30 mostra os valores de ω e χ para redes reais. Para redes com $\omega \approx 0$, como as redes *Zachary karate club*, *Jazz* e *Les Miserables* os valores de χ tende a ser alto. Redes que estão na faixa $-0.5 \leq \omega \leq 0.5$ como as redes *C. Elegans metabolic*, *US Airports* e *A song of ice and fire*, o valor de χ também tende a ser elevado. Opostamente, para redes com maior aleatoriedade, como a rede *Word adjacency*, χ tende a valores baixos e novamente isso ocorre, segundo a Tabela 30. A rede *C. Elegans metabolic* é mais “pequeno-mundo” do que a rede *C. Elegans neural*, segundo o índice ω , então χ tenderia a ser maior para a primeira rede do que para a segunda, como de fato ocorre.

Tabela 30 – Valores de acurácia do método DTWB combinando as regras B-R1 e B-R2 e para seis valores de memória combinados $\vec{\mu} = [1, 2, 3, 4, 5, 6]$

Rede	C	ω	$\chi(\mu = 1)$	$\chi(\mu = 2)$	$\chi(\mu = 3)$	$\chi(\mu = 4)$
Karate club	0,57	-0,08	0,63	0,47	0,60	0,67
Jazz	0,61	- 0,01	0,68	0,87	0,91	0,69
Les Miserables	0,57	0,03	0,55	0,36	0,50	0,58
US Airport	0,50	0,16	0,56	0,32	0,41	0,54
<i>C. elegans</i> metabolic	0,64	0,18	0,58	0,25	0,39	0,45
A song of ice and fire	0,48	0,25	0,52	0,25	0,25	0,35
<i>C. Elegans</i> neural	0,29	0,57	0,31	0,19	0,22	0,24
Word Adjacency	0,17	0,70	0,16	0,07	0,11	0,09

Fonte: Elaborada pelo autor. Redes extraídas do Konect project (24).

Para o valor de memória $\mu = 1$ o valor de χ reproduz o valor do próprio coeficiente de aglomeração da rede, uma vez que, as trajetórias para $\mu = 1$ são mais “bem comportadas” e tendem a ser menos complexas, atingindo o valor máximo dentro do intervalo $[10^{-2}, 2 \times 10^{-1}]$ e estabilizando a partir de então, como mostra a Figura 20. Já memórias elevadas, geram trajetórias mais complexas, que pode variar bastante de rede para rede, portanto, o valor de χ tende a destoar do valor de C , embora mantenha o padrão de máximo para redes de pequeno-mundo e mínimo para redes aleatórias e regulares.

A maior vantagem da métrica χ com relação à métrica ω é a velocidade. Executar o método da caminhada do turista para calcular o coeficiente χ mostrou-se mais rápido do que o cálculo de ω , especialmente para redes grandes, como mostra a Figura 22.(b). Para redes WS de mesmo tamanho, porém com probabilidades p diferentes, os dois métodos executam o cálculo em tempos semelhantes, como mostra a Figura 22.(a), sendo o tempo de execução do índice ω praticamente constante para todos os valores de p . Nessa situação o método DTW é desvantajoso com relação ao ω para memórias elevadas, no entanto, não há a necessidade de calcular χ para memórias elevadas, quando o coeficiente mostrou bons resultados para $\mu = 1$, $\mu = 2$ e $\mu = 3$, cujos tempos de execução foram da mesma ordem

do tempo de execução de ω . Já para redes grandes, o cálculo de χ mostrou-se mais rápido do que o cálculo de ω .

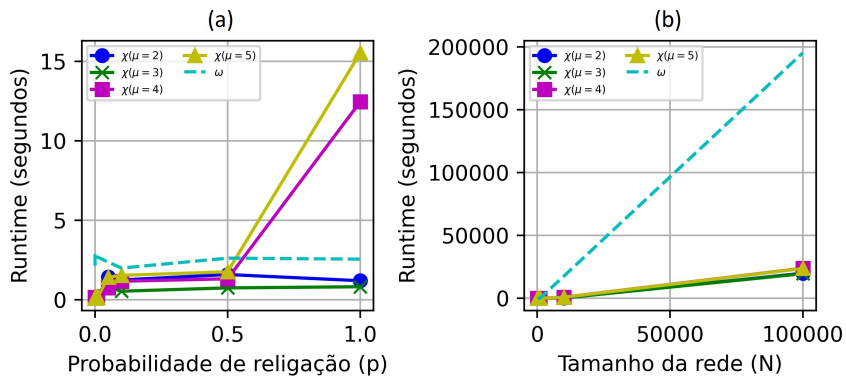


Figura 22 – Comparação de tempo de execução (em segundos) entre a métrica χ e a métrica ω . (a) Comparação de tempo de execução para redes de mesmo tamanho n e mesmo grau médio $\langle K \rangle$, mas com probabilidade de religação variando. (b) Comparação de tempo de execução para redes com mesma probabilidade de religação $p = 0.25$, mas com tamanhos (n) diferentes.

Fonte: Elaborada pelo autor.

7 CONSIDERAÇÕES FINAIS

O foco principal desse projeto foi o de desenvolver um método que permitisse o reconhecimento de padrões em redes. Dois métodos foram desenvolvidos ao longo desse projeto. O primeiro trata-se de um método específico para analisar redes de pequeno-mundo, tanto para estudar a transição entre rede regular e rede de pequeno-mundo e de rede de pequeno-mundo para rede aleatória em grafos de Watts-Strogatz, quanto para fornecer uma métrica que qualifique o quão pequeno-mundo é uma rede do mundo real. O segundo método trata-se das caminhadas determinísticas do turista com bifurcações, um método que corrige uma das falhas de seu antecessor, o descarte de informação quando há empates a partir da aplicação de uma dada regra de caminhada.

Embora ambos os métodos sejam evoluções de um método que existe há quase duas décadas, a caminhada determinística do turista, os dois métodos mostraram resultados que foram além de seu antecessor. O primeiro foi bem-sucedido em apontar a transição de comportamento regular para comportamento aleatório em redes WS e a métrica derivada deste método apresentou resultados contundentes com outras métricas bem conhecidas, bem como se mostrou computacionalmente pouco custosa. Já o DTWB mostrou bons resultados para a classificação de redes sintéticas e reais, superando seu antecessor, o DTW tradicional e superando as medidas estruturais na maior parte dos experimentos, sendo que o uso de medidas estruturais é uma das ferramentas mais tradicionais e bem-sucedidas para o reconhecimento de padrões em redes complexas. O método também conseguiu suplantar o *Life-like network automata* em alguns experimentos, um método que obteve excelentes resultados, como mostram os artigos referenciados.

Há ainda um universo de possibilidades para os dois métodos, especialmente o DTWB, onde mais regras de caminhada podem ser testadas e outras propriedades das redes podem ser exploradas pelo autômato.

Os campos de pesquisa abordados neste trabalho, a área de reconhecimento de padrão e a ciência de redes, tem sido cada vez mais importante em diversas áreas acadêmicas, da indústria e cada vez mais penetrando na vida cotidiana. A cada dia surgem novos métodos e novas aplicações para essas áreas e, sem dúvida, compreender redes e os padrões presentes nelas é compreender o próprio mundo moderno, um mundo cada vez mais conectado.

REFERÊNCIAS

- 1 THEODORIDIS, S.; KOUTROUMBAS, K. **Pattern recognition**. Burlington: Elsevier, 2006.
- 2 PAULUS, D. W.; HORNEGGER, J. **Applied pattern recognition: a practical introduction to image and speech processing in c++**. Berlin: Springer, 1998.
- 3 NEWMAN, M. E.; BARABÁSI, A.-L. E.; WATTS, D. J. **The structure and dynamics of networks**. Princeton: Princeton university press, 2006.
- 4 PAOLETTI, T. **Leonard Euler’s solution to the Königsberg Bridge problem**. Mathematical Association of America. Disponível em: <https://www.maa.org/press/periodicals/convergence/leonard-eulers-solution-to-the-konigsberg-bridge-problem>. Acesso em: 03 Mar. 2023.
- 5 SOLOMONOFF, R.; RAPOPORT, A. Connectivity of random nets. **The Bulletin of Mathematical Biophysics**, Springer, v. 13, n. 2, p. 107–117, 1951.
- 6 ERDOS, P.; RÉNYI, A. *et al.* On the evolution of random graphs. **Publications of the Mathematical Institute of the Hungarian Academy of Science**, Citeseer, v. 5, n. 1, p. 17–60, 1960.
- 7 GILBERT, E. N. Random graphs. **The Annals of Mathematical Statistics**, JSTOR, v. 30, n. 4, p. 1141–1144, 1959.
- 8 BACKES, A. R. *et al.* Deterministic tourist walks as an image analysis methodology based. *In: IBEROAMERICAN CONGRESS IN PATTERN RECOGNITION*, 11th, Cancun, Mexico, Novembro 14-17, 2006. **Proceedings[...]** (Lecture Notes in Computer Science, 4225, p. 784-793).
- 9 YANG, S. **Networks: an introduction** by MEJ Newman. Oxford, UK: Oxford University Press., 2013. 720 p.
- 10 COSTA, L. d. F. *et al.* Characterization of complex networks: a survey of measurements. **Advances in Physics**, Taylor & Francis, v. 56, n. 1, p. 167–242, 2007.
- 11 TRAVENÇOLO, B. A. N.; COSTA, L. d. F. Accessibility in complex networks. **Physics Letters A**, Elsevier, v. 373, n. 1, p. 89–95, 2008.
- 12 NEWMAN, M. E. The structure of scientific collaboration networks. **Proceedings of the National Academy of Sciences**, National Academy of Sciences, v. 98, n. 2, p. 404–409, 2001.
- 13 ALBERT, R.; BARABÁSI, A.-L. Statistical mechanics of complex networks. **Reviews of Modern Physics**, APS, v. 74, n. 1, p. 47, 2002.
- 14 MILGRAM, S. The small world problem. **Psychology Today**, New York, v. 2, n. 1, p. 60–67, 1967.
- 15 WATTS, D. J. **Six degrees: the science of a connected age**. New York: WW Norton & Company, 2004.

- 16 WATTS, D. J.; STROGATZ, S. H. Collective dynamics of ‘small-world’ networks. **Nature**, Nature Publishing Group, v. 393, n. 6684, p. 440–442, 1998.
- 17 WALSH, T. *et al.* Search in a small world. *In*: CITESEER. **Ijcai**. v. 99, p. 1172–1177, 1999.
- 18 TELESFORD, Q. K. *et al.* The ubiquity of small-world networks. **Brain Connectivity**, v. 1, n. 5, p. 367–375, Dec. 2011.
- 19 PRICE, D. J. D. S. Networks of scientific papers: the pattern of bibliographic references indicates the nature of the scientific research front. **Science**, American Association for the Advancement of Science, v. 149, n. 3683, p. 510–515, 1965.
- 20 ALBERT, R.; JEONG, H.; BARABÁSI, A.-L. Diameter of the world-wide web. **Nature**, Nature Publishing Group, v. 401, n. 6749, p. 130–131, 1999.
- 21 FALOUTSOS, M.; FALOUTSOS, P.; FALOUTSOS, C. On power-law relationships of the internet topology. **ACM SIGCOMM Computer Communication Review**, ACM, v. 29, n. 4, p. 251–262, 1999.
- 22 DOROGOVTSEV, S. N.; MENDES, J. F. F. Evolution of networks with aging of sites. **Physical Review E**, APS, v. 62, n. 2, p. 1842, 2000.
- 23 KRAPIVSKY, P. L.; REDNER, S.; LEYVRAZ, F. Connectivity of growing random networks. **Physical Review Letters**, APS, v. 85, n. 21, p. 4629, 2000.
- 24 KUNEGIS, J. Konect: the koblenz network collection. *In*: INTERNATIONAL CONFERENCE ON WORLD WIDE WEB, 22., May 2013. Rio de Janeiro. **Proceedings[...]**. New York: Association for Computing Machinery, 2013. p. 1343–1350.
- 25 COMIN, C.; COSTA, L. d. F. **Spatial Networks**: when topology meets geometry (CDT-3). 2018. DOI:10.13140/RG.2.2.11110.32322.
- 26 DALL, J.; CHRISTENSEN, M. Random geometric graphs. **Physical Review E**, APS, v. 66, n. 1, p. 016121, 2002.
- 27 BARTHÉLEMY, M. Spatial networks. **Physics Reports**, Elsevier, v. 499, n. 1-3, p. 1–101, 2011.
- 28 WAXMAN, B. M. Routing of multipoint connections. **IEEE Journal on Selected Areas in Communications**, IEEE, v. 6, n. 9, p. 1617–1622, 1988.
- 29 RENDELL, P. A universal turing machine in conway’s game of life. *In*: INTERNATIONAL CONFERENCE ON HIGH PERFORMANCE COMPUTING & SIMULATION, 2011, Istambul. **Proceedings[...]**. Istambul: IEEE, 2011. p. 764–772, 2011.
- 30 MIRANDA, G. H. B.; MACHICAO, J.; BRUNO, O. M. Exploring spatio-temporal dynamics of cellular automata for pattern recognition in networks. **Scientific Reports**, Nature Publishing Group, v. 6, n. 1, p. 1–15, 2016.
- 31 RIBAS, L. C.; MACHICAO, J.; BRUNO, O. M. Life-like network automata descriptor based on binary patterns for network classification. **Information Sciences**, Elsevier, v. 515, p. 156–168, 2020.

-
- 32 LOVÁSZ, L. Random walks on graphs. **Combinatorics**, Paul erdos is eighty, v. 2, n. 4, p. 1–46, 1993.
- 33 KINOUCI, O. *et al.* Deterministic walks in random networks: an application to thesaurus graphs. **Physica A: statistical mechanics and its applications**, Elsevier, v. 315, n. 3-4, p. 665–676, 2002.
- 34 BACKES, A. R. *et al.* Texture analysis and classification using deterministic tourist walk. **Pattern Recognition**, Elsevier, v. 43, n. 3, p. 685–694, 2010.
- 35 BACKES, A. R.; MARTINEZ, A. S.; BRUNO, O. M. Texture analysis using graphs generated by deterministic partially self-avoiding walks. **Pattern Recognition**, Elsevier, v. 44, n. 8, p. 1684–1689, 2011.
- 36 RIBAS, L. C.; BRUNO, O. M. **Fast deterministic tourist walk for texture analysis**. Disponível em: <https://arxiv.org/abs/1611.08624>. Acesso em: 03 Mar. 2023.
- 37 GONÇALVES, W. N. *et al.* Texture descriptor based on partially self-avoiding deterministic walker on networks. **Expert Systems with Applications**, Elsevier, v. 39, n. 15, p. 11818–11829, 2012.
- 38 WEBER, A. G. **The USC-SIPI image database: version 5**. Disponível em: <http://www.kaggle/datasets/luffyluffyluffy/the-uscsipi-image-database>. Acesso em: 03 Mar. 2023.
- 39 GONÇALVES, W. N.; MARTINEZ, A. S.; BRUNO, O. M. Complex network classification using partially self-avoiding deterministic walks. **Chaos: an interdisciplinary journal of nonlinear science**, American Institute of Physics, v. 22, n. 3, p. 033139, 2012.
- 40 MERENDA, J. V.; BRUNO, O. M. Using deterministic self-avoiding walks as a small-world metric on watts–strogatz networks. **Physica A: statistical mechanics and its applications**, Elsevier, v. 621, p. 128713, 2023.
- 41 KANEHISA, M.; GOTO, S. Kegg: Kyoto encyclopedia of genes and genomes. **Nucleic Acids Research**, Oxford University Press, v. 28, n. 1, p. 27–30, 2000.
- 42 LESKOVEC, J.; KREVL, A. **SNAP: Stanford network analysis project**. 2014. Disponível em: <http://snap.stanford.edu/>. Acesso em: 06 Mar. 2023.
- 43 FISHER, R. A. The use of multiple measurements in taxonomic problems. **Annals of Eugenics**, Wiley Online Library, v. 7, n. 2, p. 179–188, 1936.
- 44 MCLACHLAN, G. J. **Discriminant analysis and statistical pattern recognition**. New York: John Wiley & Sons, 2005.
- 45 GEWERS, F. L. *et al.* Principal component analysis: A natural approach to data exploration. **ACM Computing Surveys**, ACM New York, NY, USA, v. 54, n. 4, p. 1–34, 2021.

APÊNDICES

APÊNDICE A – MÉTODOS DE CLASSIFICAÇÃO E VALIDAÇÃO: LDA, PCA E LOOCV

A.1 Linear Discriminant Analysis

O *Linear discriminant analysis* (LDA) é um método supervisionado de aprendizado de máquinas para a classificação de objetos em categorias ou classes. O método LDA, assim como o PCA, pode ser usado para classificação, redução de dimensionalidade ou visualização de dados. Este método usa análise estatística, baseada no teorema de Bayes, para modelar as diferenças interclasses. (43,44)

Seja C o número de classes. O LDA cria C regiões disjuntas para representar cada classe k baseado em funções de densidade de probabilidade ($f_k(x)$). Em geral, o método LDA assume uma distribuição normal dos dados, embora o método também se aplique a dados que não estão normalmente distribuídos. Outra suposição do método é de que as classes compartilham a mesma variância.

Seja $\vec{X} = [x_1, x_2, \dots, x_n]$ um vetor de característica pertencente a classe k . Seja $f_k(\vec{X})$ a função densidade de probabilidade (normal) associada a classe k . Como o vetor \vec{X} possui dimensão maior do que 1, temos de assumir uma média para cada dimensão, formando um vetor de médias $\vec{\mu}_k = [\mu_{k,1}, \mu_{k,2}, \dots, \mu_{k,n}]$ que define a média da classe k . Seja Σ a matriz covariância a qual mede a covariância entre as componentes do vetor \vec{X} e tal essa matriz é comum a todas as classes. A função densidade de probabilidade para a classe k é definida como:

$$f_k(\vec{X}) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\vec{X} - \vec{\mu}_k)^T \Sigma^{-1} (\vec{X} - \vec{\mu}_k)\right) \quad (\text{A.1})$$

Seja um *dataset* com m vetores de característica \vec{X} , com n características cada. A probabilidade de um vetor \vec{X} pertencer a k -ésima classe é dada pelo teorema de Bayes:

$$P(\vec{X}|y = k) = \frac{\Pi_k \cdot f_k(\vec{X})}{\sum_{i=1}^C \Pi_i \cdot f_i(\vec{X})} \quad (\text{A.2})$$

Seja m_k a quantidade de vetores no *dataset* que são pertencentes a k -ésima classe, então definimos a razão Π_k de um dado vetor \vec{X} como: $\Pi_k = m_k/m$.

Como o método LDA é um método de aprendizado de máquinas supervisionado, é necessário oferecer um conjunto de treino ao método. Com o treino serão estimados o vetor de média de cada classe, a probabilidade Π_k de cada classe e a variância.

Dado um vetor \vec{X} de teste, o método, agora calibrado pelo conjunto treino, conseguirá prever a classe dessa nova entrada através da função discriminante δ_k :

$$\delta_k = \vec{X}^T \Sigma^{-1} \vec{\mu}_k - \frac{1}{2} \vec{\mu}_k^T \Sigma^{-1} \vec{\mu}_k + \ln(\Pi_k) \quad (\text{A.3})$$

Cada classe terá sua função discriminante a qual é linear com respeito ao \vec{X} e em razão disso o método recebe tal nome. O LDA escolhe o valor mais alto dentre os δ_k e então a classe k associada a esse valor é selecionada para o vetor de entrada.

A.2 Principal Component Analysis

O método *Principal component analysis* (PCA) é um método não-supervisionado de aprendizado de máquinas usado como um classificador, mas também como forma de redução de dimensionalidade e visualização de dados. (45) O método consiste em uma transformação linear que rotaciona o espaço para ajustá-lo aos dados de entrada, onde o primeiro eixo, ou primeiro eixo principal, é a direção de maior dispersão dos dados e o segundo eixo, ou segundo eixo principal, representa a direção da segunda maior dispersão dos dados. Embora o PCA pode ter d eixos principais, com $d \leq n$, em geral, consideramos os dois principais eixos, para possibilitar a visualização dos dados em um gráfico de dispersão. O método se ajusta aos dados de tal forma que os dois principais eixos concentram quase toda a informação contida nos dados, portanto, não convêm utilizar mais eixos.

Seja \vec{X}' o *dataset* de entrada com m vetores de característica (linhas) e n atributos (colunas). Para facilitar a notação, tomarei a transposta desse vetor: $\vec{X} = \vec{X}'^T$. Agora os atributos são as linhas e cada vetor compõe uma coluna de \vec{X} . Precisamos encontrar uma matriz M que rotacione \vec{X} de forma a concentrar a maioria da informação, antes contida em um espaço n -dimensional, em um espaço reduzido, um espaço bidimensional.

Seja $\vec{\mu}_i = [\mu_{i,1}, \mu_{i,2}, \dots, \mu_{i,n}]^T$ o vetor média associado a média de cada atributo, com cada componente sendo definida como:

$$\mu_i = \frac{1}{m} \sum_{j=1}^m X_{ij} \quad (\text{A.4})$$

Podemos, para facilitar as contas, deslocar o conjunto de dados para a origem do espaço:

$$\hat{X}_i = \vec{X}_i - \mu_i \quad (\text{A.5})$$

Seja $cov(\hat{X})$ a matriz covariância associada ao *dataset* \hat{X} e definida como:

$$cov(\hat{X}) = \frac{1}{m-1} \hat{X} \hat{X}^T \quad (\text{A.6})$$

Tomando todos os autovalores não-negativos λ_i de $cov(\hat{X})$, e seus respectivos autovetores v_i , construímos a matriz da transformação:

$$\hat{Y} = M\hat{X} \quad (\text{A.7})$$

com M definido como:

$$M = \begin{bmatrix} \vec{v}_1 \\ \vec{v}_2 \\ \vdots \\ \vec{v}_m \end{bmatrix} \quad (\text{A.8})$$

Note que a dimensionalidade de \hat{Y} é a mesma de \hat{X} e está apropriado para tarefas de classificação. No entanto, se o objetivo for a visualização, a redução de dimensionalidade é necessária, entretanto o método do PCA já oferece essa possibilidade, basta usar os dois principais autovetores e descartar o restante, sendo assim, a dimensionalidade do espaço é reduzida para dois sem perda significativa de informação.

A.3 Método Leave-One-Out cross-validation

O método *Leave-one-out* (LOOCV) é um método de validação cruzada responsável por validar a classificação efetuada por um algoritmo classificador.

Seja um *dataset* $X = [\vec{X}_1, \vec{X}_2, \vec{X}_3, \dots, \vec{X}_m]$ com m vetores de característica, cada um com n componentes. O método LOOCV divide o *dataset* tomando $m - 1$ vetores para formar o conjunto treino e o vetor restante é usado como teste.

Na primeira rodada o subconjunto $[\vec{X}_2, \vec{X}_3, \dots, \vec{X}_m]$ é selecionado como conjunto treino e o subconjunto $[\vec{X}_1]$ é selecionado como conjunto teste. Na segunda rodada o conjunto $[\vec{X}_1, \vec{X}_3, \dots, \vec{X}_m]$ é selecionado como conjunto treino e $[\vec{X}_2]$ é selecionado como conjunto teste. Na terceira rodada, $[\vec{X}_1, \vec{X}_2, \dots, \vec{X}_m]$ é o treino e $[\vec{X}_3]$ é o teste. O método repete o mesmo passo por m rodadas, sempre deixando um vetor de fora o qual será usado como teste. A cada rodada o método avalia o desempenho do classificador selecionado (*e.g.*, LDA) de acordo com alguma métrica. As métricas de avaliação mais utilizadas são a acurácia e o erro quadrático médio (MSE).