
UNIVERSIDADE ESTADUAL DE MARINGÁ
DEPARTAMENTO DE FÍSICA

ARTHUR AUGUSTO BARIZON PESSA

ANÁLISE DE SÉRIES TEMPORAIS VIA
REDES ORDINAIS

Maringá, Novembro de 2019.

UNIVERSIDADE ESTADUAL DE MARINGÁ
DEPARTAMENTO DE FÍSICA

ARTHUR AUGUSTO BARIZON PESSA

ANÁLISE DE SÉRIES TEMPORAIS VIA
REDES ORDINAIS

Dissertação apresentada ao Programa de Pós-Graduação em Física da Universidade Estadual de Maringá como requisito parcial para obtenção do título de Mestre em Física.

Orientador: Prof. Dr. Haroldo Valentin Ribeiro

Maringá, Novembro de 2019.

Dados Internacionais de Catalogação-na-Publicação (CIP)
(Biblioteca Central - UEM, Maringá - PR, Brasil)

B557a

Pessa, Arthur Augusto Barizon

Análise de séries temporais via redes ordinais / Arthur Augusto Barizon Pessa. --
Maringá, PR, 2020.

86 f.color., figs., tabs.

Orientador: Prof. Dr. Haroldo Valentin Ribeiro.

Dissertação (Mestrado) - Universidade Estadual de Maringá, Centro de Ciências
Exatas, Departamento de Física, Programa de Pós-Graduação em Física, 2020.

1. Redes complexas. 2. Redes ordinais. 3. Séries temporais. 4. Processos estocásticos.
5. Sistemas dinâmicos. I. Ribeiro, Haroldo Valentin, orient. II. Universidade Estadual de
Maringá. Centro de Ciências Exatas. Departamento de Física. Programa de Pós-
Graduação em Física. III. Título.

CDD 23.ed. 530.475

Resumo

O estudo de redes complexas tem experimentado imenso sucesso nas últimas duas décadas. Além de serem usadas como representação simplificada de diversos sistemas complexos, as redes têm sido usadas como ferramenta de investigação de aspectos dinâmicos subjacentes a séries temporais de natureza não linear ou estocástica. Essa conexão entre a análise de séries temporais e redes emergiu a partir de diversos algoritmos para a transformação dessas sequências numéricas em grafos. Apesar desses algoritmos divergirem entre si quanto ao significado dos vértices e arestas da rede, todos compartilham o objetivo comum de extrair informação de uma série temporal a partir da topologia da rede resultante do mapeamento. Nesta dissertação, revisitamos alguns aspectos dos principais algoritmos usados no mapeamento de séries temporais (grafos de visibilidade, grafos de quantil e redes de recorrência) e investigamos cuidadosamente uma transformação particular, as chamadas redes ordinais. Esse mapeamento tem sido muito utilizado no contexto de séries temporais não lineares e muito pouco é conhecido sobre suas características estruturais quando aplicado em séries temporais simples ou estocásticas. Neste trabalho, contribuímos para diminuir essa lacuna por meio da investigação de propriedades topológicas de redes ordinais obtidas de séries temporais periódicas, aleatórias e amostras de movimento browniano fracionário. Além disso, mostramos que redes ordinais podem ser usadas para estimar o expoente de Hurst de séries de movimento browniano fracionário e identificar mudanças sutis em séries temporais de atividade sísmica terrestre causadas por grandes terremotos.

Palavras-chave: Redes complexas. Redes ordinais. Séries temporais. Processos estocásticos. Sistemas dinâmicos. Terremotos.

Abstract

Complex networks have been immensely successful over the past two decades. In addition to being used as simplified representations of several complex systems, networks have been used as a tool for investigating dynamical features underlying time series of stochastic or chaotic nature. This connection between time series analysis and networks has emerged through algorithms which transform such numerical sequences into graphs. Despite divergent definitions of vertices and edges, all these algorithms share the common goal of extracting information from a time series using the topology of the resulting network. In this dissertation, we revisit some of the main algorithms used for mapping time series into graphs (visibility graphs, quantile graphs and recurrence networks) and carefully investigate a particular type of transformation called ordinal networks. This transformation has been mainly used in the context of nonlinear time series and little is known about its structural characteristics when applied to simple or stochastic time series. In this work, we contribute with results regarding topological features of ordinal networks mapped from periodic, random, and fractional Brownian motion time series. We further show that ordinal networks can be used to estimate the Hurst exponent of fractional Brownian motion samples and identify subtle changes in Earth's seismic activity caused by large earthquakes.

Keywords: Complex networks. Ordinal networks. Time series. Stochastic processes. Dynamical systems. Earthquakes.

Introdução	7
1 Transformações de séries temporais em redes complexas	10
1.1 Grafos de visibilidade	10
1.1.1 Grafos de visibilidade e o expoente de Hurst	14
1.1.2 Ruído, caos e grafos de visibilidade horizontal	15
1.2 Grafos de quantil	18
1.2.1 Grafos de quantil e o expoente de Hurst	20
1.3 Redes de recorrência	20
1.3.1 Redes de recorrência aplicadas ao mapa logístico	23
1.4 Redes Ordinais	25
1.4.1 Redes ordinais e o sistema de Rössler	28
2 Propriedades básicas e aplicações de redes ordinais a processos estocásticos	32
2.1 Estrutura e propriedades de redes ordinais	32
2.1.1 Topologia de redes ordinais de séries temporais simples	34
2.1.2 Redes ordinais aleatórias	35
2.2 Sinal, ruído e redes ordinais aleatórias	38
2.3 Redes ordinais e correlações em séries temporais	43
2.4 Redes ordinais aplicadas a séries de atividade sísmica	48
Conclusões e Perspectivas	50
A Redes complexas	52
A.1 Redes complexas e sua representação	52

A.2 Modelos de redes complexas	54
B Movimento browniano fracionário	58
B.1 Séries temporais, autocorrelações e processos estocásticos	58
B.2 Movimento browniano fracionário	62
B.3 Simulação do movimento browniano fracionário	63
C Métodos de aprendizagem estatística	73
C.1 Aprendizagem estatística e algoritmo k -primeiros vizinhos	73
C.2 Aplicação do algoritmo de k -primeiros vizinhos à estimativa do expoente de Hurst	75
D Análise de flutuação destendenciada	78

O estudo de redes complexas teve seu início na abordagem desenvolvida por Leonhard Euler no tratamento do problema das sete pontes de Königsberg, no início do século XVIII [1, 2]. Nessa abordagem, Euler representava porções de terra dessa cidade como vértices e pontes que ligavam essas porções como arestas do que hoje chamaríamos de um grafo [1,2]. Apesar desse precedente histórico, foi apenas nas últimas décadas que redes complexas começaram a ser largamente aplicadas na representação e caracterização de diversos sistemas [1–4]. Esse sucesso está diretamente relacionado à simplicidade com a qual virtualmente qualquer sistema interagente pode ser representado por conjuntos de vértices e arestas [2,4].

Para além de uma mera representação sintética de um sistema complexo, um dos últimos desenvolvimentos no estudo de redes aponta em uma direção de maior abstração, culminando em algoritmos de mapeamento de séries temporais em grafos e dando origem às chamadas redes de séries temporais [5]. Os primeiros trabalhos que usam de grafos na análise de séries temporais datam de pouco mais de dez anos e apresentam uma importante conexão com o estudo de sistemas dinâmicos [5,6]. De um modo geral, as propostas para o mapeamento de uma série temporal em uma rede encontradas na literatura podem ser classificadas em três categorias [5]: grafos de visibilidade, redes de transição e redes de proximidade. Nas diversas variantes de grafos de visibilidade, cada observação de uma série temporal corresponde a um vértice na rede e pares de vértices são conectados caso as observações associadas a esses vértices satisfaçam uma “condição de visibilidade” [7, 8]. Em redes de transição, os vértices da rede correspondem a um conjunto finito de estados associados a segmentos de uma série temporal e arestas são traçadas entre vértices a partir das sucessões de estados que ocorrem ao longo da série [9,10]. Em redes de proximidade, os vértices da rede podem representar um conjunto de observações da série [11] ou um conjunto finito de estados [12], semelhantemente às redes de transição. As arestas, por sua vez, ligam vértices (pares de conjuntos de observações ou de estados) suficientemente próximos (semelhantes) entre si de

acordo com uma medida de distância [11, 12].

Esses diferentes procedimentos de mapeamento de séries temporais em redes têm grande potencial para contribuir com métodos novos e mais eficientes para a análise de dados. Trata-se de uma contribuição importante tendo em vista a crescente quantidade de dados envolvidos em pesquisas das mais variadas origens [13]. De fato, não apenas o volume de dados tem aumentado, mas também a riqueza de detalhes em muitos estudos não encontra precedentes. Até mesmo a investigação de objetos relativamente mais simples como séries temporais pode se mostrar desafiadora dependendo do contexto do problema e da quantidade de dados envolvidos na análise. Nesse contexto, as abordagens relacionadas às redes de séries temporais representam uma contribuição relevante. Ao conectar séries temporais com redes complexas, todo o aparato teórico de redes complexas fica disponível à análise de séries temporais. Aqui, vale ressaltar a possibilidade do caminho inverso, ou seja, o mapeamento de redes complexas em séries temporais [9, 14], que embora tenha sido pouco explorado na caracterização de redes complexas, permite imaginar uma espécie de dualidade entre redes e séries temporais [9].

Conforme já mencionamos, algoritmos para a construção de redes de séries temporais surgiram a pouco na literatura científica. Entre esses, as redes ordinais, um exemplo de rede de transição, estão entre os mais recentes. Desde seu surgimento na literatura [10], redes ordinais têm sido usadas principalmente para investigar séries temporais de sistemas dinâmicos não lineares (como mapas caóticos) e poucos são os trabalhos que se valeram dessa abordagem para analisar séries temporais de natureza estocástica ou séries empíricas. De fato, apesar de existirem algumas generalizações desse algoritmo [10, 15–18], muito pouco se conhece acerca da estrutura básica dessas redes. Esta dissertação tem como principal objetivo contribuir para o desenvolvimento do formalismo das redes ordinais, especialmente no que diz respeito à sua utilização no estudo de séries temporais estocásticas [19]. Com esse fim, este trabalho encontra-se organizado como segue.

No Capítulo 1, apresentamos uma breve revisão de alguns dos mais importantes desenvolvimentos encontrados na literatura relativos a redes de séries temporais. Para tanto, descrevemos os principais algoritmos que mapeiam séries temporais em redes: grafos de visibilidade [7, 8], grafos de quantil [9], redes de recorrência [12] e redes ordinais [10]. Além disso, reproduzimos e discutimos algumas aplicações desses métodos a séries caóticas e estocásticas. No Capítulo 2, apresentamos alguns resultados (até então desconhecidos) acerca da topologia de rede ordinais mapeadas de sinais simples e séries temporais de natureza estocástica. Usando esses resultados, empregamos as redes ordinais na tarefa de detecção de comportamento aleatório em séries temporais curtas. Além disso, mostramos que é possível usar propriedades de redes ordinais para identificar o expoente de Hurst em amostras de movimento browniano fracionário com grande precisão. Finalmente, investigamos mudanças em métricas de redes ordinais de séries de atividade sísmica terrestre causadas pela ocorrência

de grandes terremotos.

Por fim, encerramos este trabalho com considerações finais acerca dos tópicos apresentados e expomos nossas perspectivas de futuras investigações. Na tentativa de tornar a leitura deste trabalho mais objetiva, apresentamos na forma de apêndices alguns aspectos relacionados à teoria de redes complexas, movimento browniano fracionário, métodos de aprendizagem estatística e análise de flutuação destendenciada.

Transformações de séries temporais em redes complexas

Neste primeiro capítulo, apresentamos quatro diferentes transformações de séries temporais em redes complexas: grafos de visibilidade, grafos de quantil, redes de recorrência e redes ordinais. O procedimento de construção de cada rede é ilustrado com exemplos simples. Desenvolvemos, ainda, uma aplicação para cada método apresentado e comparamos nossos resultados com aqueles reportados na literatura.

1.1 Grafos de visibilidade

O algoritmo de transformação de uma série temporal em um grafo de visibilidade foi inicialmente proposto por Lacasa *et al.* [7]. O conceito de visibilidade entre elementos de uma série é como o algoritmo extrai informação de sequências numéricas, estabelece conexões e define a estrutura da rede complexa. Nessa proposta de construção de grafos, cada observação em uma série temporal corresponde a um vértice na respectiva rede complexa. A relação entre os vértices, que é capturada pelas arestas, é o que chamamos de relação ou condição de visibilidade. Caso dois valores quaisquer pertencentes a uma série sejam “visíveis” um ao outro, existe uma aresta conectando seus respectivos vértices no grafo de visibilidade.

De modo a formalizar o conceito de visibilidade, passamos a considerar uma série temporal denotada por $\{x_t\}_{t=1,\dots,N}$, sendo $x_i = x(t_i)$ uma observação ou medição realizada ao tempo t_i . Mais especificamente, consideramos dois valores x_i e x_k dessa sequência com $t_i < t_k$. Em seguida, encontramos a reta $y(t)$ que liga o ponto (t_i, x_i) ao ponto (t_k, x_k) . Naturalmente, podemos encontrar os valores $y(t_j)$ que essa reta assume nos instantes de tempo t_j intermediários ao intervalo (t_i, t_k) . Caso nenhum x_j (com $i < j < k$) da série temporal seja maior ou igual que o valor correspondente de $y(t_j)$, x_i e x_k são visíveis um ao outro.

Podemos expressar essa condição de visibilidade na forma da desigualdade

$$x_j < x_k + (x_i - x_j) \left(\frac{t_j - t_k}{t_j - t_i} \right) \quad \forall \quad t_i < t_j < t_k, \quad (1.1)$$

na qual o lado direito nada mais é do que o valor de $y(t_j)$. Se a condição de visibilidade expressa pela Eq. 1.1 é satisfeita, os vértices correspondentes aos elementos x_i e x_k são conectados na rede por uma aresta. Ainda mais, como a relação de visibilidade é uma relação recíproca, essas ligações são não direcionadas¹.

Para esclarecer melhor a desigualdade anterior, notamos que a condição de visibilidade entre dois pontos (t_i, x_i) e (t_k, x_k) representa uma restrição nas possíveis inclinações das retas que ligam (t_i, x_i) a cada um dos (t_j, x_j) , com $t_i < t_j < t_k$. Desse modo, caso a inclinação de todas as retas entre (t_i, x_i) e os possíveis (t_j, x_j) seja menor que o coeficiente angular da reta $y(t)$, x_i e x_k satisfazem o critério de visibilidade. A relação entre as inclinações é dada por

$$\frac{x_j - x_i}{t_j - t_i} < \frac{x_j - x_k}{t_j - t_k} \quad \forall \quad t_i < t_j < t_k, \quad (1.2)$$

que, rearranjada, leva à Eq. 1.1. Ao analisar o critério de visibilidade para todas as combinações de pares de elementos de uma série, obtemos o grafo de visibilidade.

A Figura 1.1 ilustra o processo de construção de um grafo de visibilidade a partir de uma série temporal simples. Podemos observar nessa figura que dois pontos consecutivos x_i e x_{i+1} são sempre conectados, visto que não existe qualquer x_j intermediário a x_i e x_{i+1} . Esse fato tem implicações diretas na estrutura dos grafos de visibilidade. Uma delas é que esse procedimento gera redes conectadas, sendo sempre possível encontrar um caminho ligando qualquer par de nós da rede [5, 7]. Além disso, esperamos que elementos excepcionalmente grandes (em valor absoluto) de uma série sejam visíveis a muitos outros, de modo que os vértices correspondentes a esses pontos devem se conectar a muitos outros vértices do grafo de visibilidade [7, 20].

Como todas as propostas para mapear séries temporais em redes complexas, os grafos de visibilidade têm por objetivo extrair informação do sinal original que gerou a rede e servir como uma ferramenta de análise da série. Um aspecto importante dessas redes diz respeito a como suas propriedades topológicas estão relacionadas a características das séries temporais. Entre outros resultados, foi observado que sequências de números aleatórios são transformadas em grafos aleatórios, enquanto séries autossimilares (fractais) dão origem a redes livres de escala e sinais periódicos produzem grafos regulares [7]. Sendo assim, os grafos de visibilidade podem capturar e revelar informação não trivial acerca da dinâmica de uma série temporal [7].

Desde o trabalho seminal de Lacasa e colaboradores [7], esse método de transformação

¹O Apêndice A traz uma breve revisão de diferentes tipos e modelos de redes complexas.

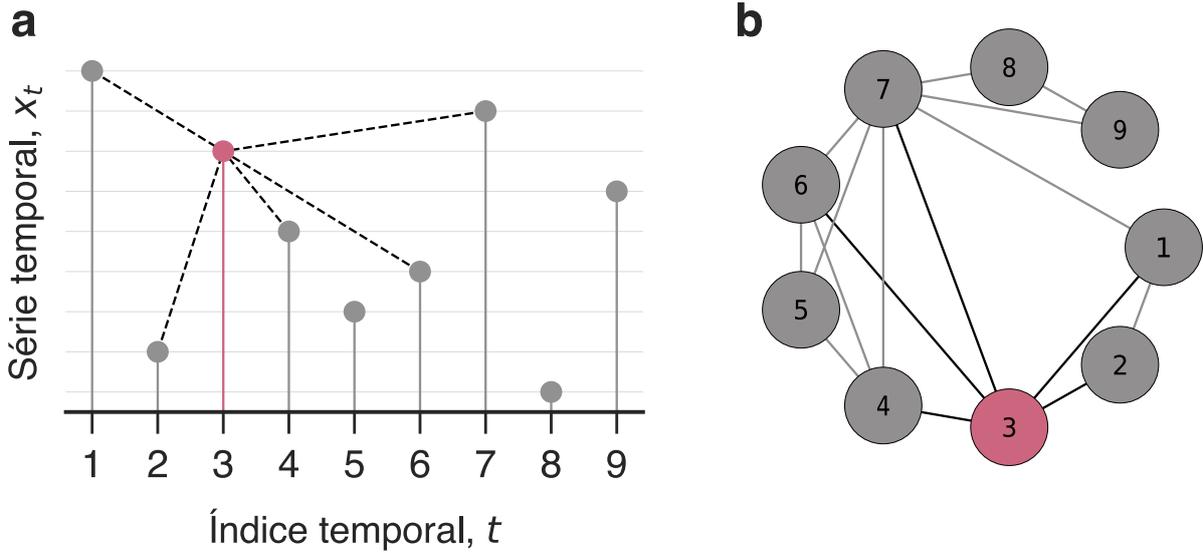


Figura 1.1: Ilustração do processo de construção de um grafo de visibilidade. (a) A série temporal $x_t = \{8, 1, 6, 4, 2, 3, 7, 0, 5\}$ é mostrada e o dado $x_3 = 6$ é destacado em vermelho. Retas tracejadas em preto indicam os pontos da série que “enxergam” x_3 ; são eles: x_1, x_2, x_4, x_6 e x_7 . (b) Grafo de visibilidade obtido a partir da série temporal x_t . O vértice correspondente a x_3 e suas conexões são destacadas nessa visualização.

de séries em redes tem sido extensamente estudado e vários outros algoritmos modificados para a construção de redes de visibilidade têm sido propostos² [5]. Uma das modificações mais importantes da proposta original é a que dá origem aos chamados grafos de visibilidade horizontal [8].

Na variante de visibilidade horizontal, ao invés do critério mais “natural” de visibilidade discutido anteriormente, adota-se uma regra mais restritiva. Especificamente, desenhada uma linha reta horizontal entre dois valores de uma série (à altura do dado de menor amplitude), esse critério de visibilidade horizontal é satisfeito caso nenhum elemento intermediário da série esteja acima dessa reta. Assim como no caso dos grafos de visibilidade, cada observação x_i de uma série $\{x_t\}_{t=1,\dots,N}$ corresponde a um nó do grafo de visibilidade horizontal. As arestas desse grafo ligam pares de nós (elementos da série) de mútua visibilidade horizontal.

Para expressar esse critério de visibilidade horizontal formalmente, sejam x_i e x_k dois pontos pertencentes a uma série $\{x_t\}_{t=1,\dots,N}$ (com $t_i < t_k$) e seja x_j tal que $t_i < t_j < t_k$. Nesse caso, a condição de visibilidade horizontal entre x_i e x_k é definida por

$$x_j < x_i, x_k \quad \forall t_i < t_j < t_k. \quad (1.3)$$

Novamente, como visibilidade horizontal é uma relação recíproca, a rede produzida é não direcionada, uma vez que dois elementos consecutivos de uma série temporal são sempre

²Em razão das muitas variações dos grafos de visibilidade que surgiram na literatura, as redes obtidas a partir do algoritmo original são também chamadas de grafos de visibilidade natural.

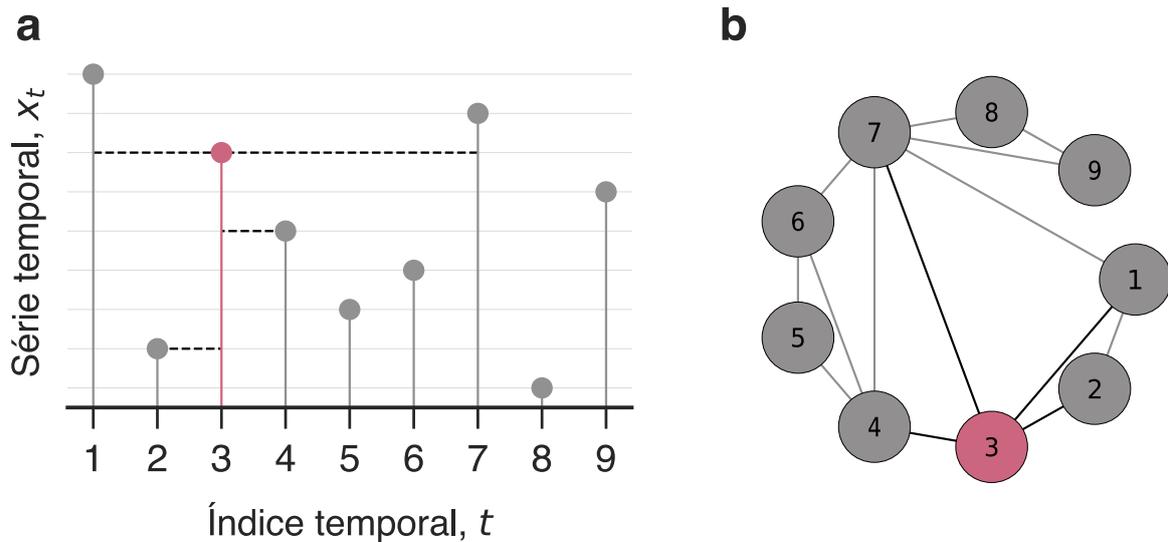


Figura 1.2: Ilustração do processo de construção de um grafo de visibilidade horizontal. (a) A série temporal $x_t = \{8, 1, 6, 4, 2, 3, 7, 0, 5\}$ é mostrada com destaque para o ponto $x_3 = 6$. As retas tracejadas em preto indicam os pontos da série que “enxergam horizontalmente” x_3 ; são eles: x_1, x_2, x_4 e x_7 . Note, por exemplo, que x_3 e x_6 não são conectados no grafo de visibilidade horizontal, diferentemente do que aconteceria no caso do grafo de visibilidade. (b) Grafo de visibilidade horizontal obtido a partir da série temporal x_t . O vértice correspondente a x_3 e suas arestas são destacadas nessa visualização.

“horizontalmente visíveis” um ao outro [8]. Notamos ainda que as arestas do grafo de visibilidade horizontal são um subconjunto das arestas do grafo de visibilidade quando ambos são obtidos a partir de uma mesma série temporal [8]. Esse resultado é uma consequência das desigualdades expressas nas Eqs. 1.1 e 1.3. O processo de construção de um grafo de visibilidade horizontal é ilustrado na Figura 1.2.

Vale ressaltar ainda que muitas outras variantes de construção de uma rede de visibilidade podem ser encontradas na literatura. Entre essas variantes, temos os grafos de visibilidade (natural ou horizontal) pesados, os quais consideram o inverso da distância entre duas observações em uma série temporal como peso da ligação entre seus respectivos nós [21]. Os grafos de visibilidade horizontal podem também ser construídos de modo que relações de causa e efeito entre os dados de uma série sejam levadas em conta. Para tanto, as arestas passam a ser direcionadas de tal maneira que caso x_i e x_j ($t_i < t_j$) sejam ligados no grafo, essa ligação é representada por uma aresta que aponta do vértice correspondente a x_i para o vértice correspondente a x_j [22]. Existem também os grafos de visibilidade natural paramétricos, que atribuem ao mesmo tempo peso e direção às arestas [23]. Nesse caso, entretanto, o peso das ligações é dado pela inclinação da reta que liga dois pontos x_i e x_j e a direção leva em conta, novamente, relações de causa e efeito. Além disso, um parâmetro adicional é introduzido nessa abordagem, trata-se de um valor limite para a inclinação da reta que une

x_i e x_j , de modo que se o coeficiente angular da reta que une esses dois pontos for maior que o ângulo máximo, os vértices correspondentes não são conectados, mesmo que o critério de visibilidade seja satisfeito.

Por fim, é importante mencionar que os grafos de visibilidade e suas variações têm sido aplicados a séries temporais das mais variadas origens e dos mais diversos campos de estudo, desde engenharia e medicina até economia e geofísica [5, 6]. Dessa forma, esses grafos constituem uma classe particularmente relevante de mapeamento de séries temporais em redes complexas.

1.1.1 Grafos de visibilidade e o expoente de Hurst

Dentre as várias aplicações dos grafos de visibilidade, destacamos seu uso no estudo de autocorrelações de longo alcance em séries temporais, tal como estabelecido no trabalho de Lacasa *et al.* [20].

A fim de analisar como a estrutura dos grafos de visibilidade depende de autocorrelações existentes em séries temporais, utilizamos uma generalização do movimento browniano usual, denominada movimento browniano fracionário [24], que corresponde a uma caminhada aleatória não markoviana na qual posições futuras do caminhante dependem de todos os seus passos anteriores³. Um movimento browniano fracionário $B_h(t)$ é um processo estocástico caracterizado por um parâmetro $h \in (0, 1)$ conhecido como expoente de Hurst. A depender do valor de h temos três regimes diferentes: o regime antipersistente ($0 < h < \frac{1}{2}$), o regime persistente ($\frac{1}{2} < h < 1$) e o passeio aleatório usual ($h = \frac{1}{2}$). Por persistência e antipersistência, referimo-nos a tendências de manutenção (por exemplo, passos para frente tendem a ser seguidos de passos para frente) ou alternância (por exemplo, passos para frente tendem a ser seguidos de passos para trás) de sinal nos incrementos do movimento browniano fracionário.

Além das autocorrelações de longo alcance entre seus incrementos, o movimento browniano fracionário produz séries temporais autossimilares. Como já mencionamos, o algoritmo de visibilidade mapeia séries fractais em redes livres de escala, o que significa que a cauda da distribuição de grau dos vértices da rede assintoticamente tem a forma de uma lei de potência $p(k) \sim k^{-\gamma}$ [25], sendo γ um expoente. Coube a Lacasa *et al.* [20] estabelecerem uma conexão entre o expoente γ e o parâmetro de Hurst h conforme passamos a descrever.

Começamos com a hipótese de que os maiores valores de uma série temporal de movimento browniano fracionário dão origem aos vértices mais conectados da rede (chamados *hubs*) e, por conseguinte, podem ser associados ao comportamento de lei de potência da distribuição de grau da rede. De fato, segundo Lacasa e colaboradores [20], denotando por $B(t_0)$ um desses pontos extremos, a probabilidade desse vértice estar conectado a k outros nós da rede

³O Apêndice B apresenta mais detalhes sobre correlações em séries temporais e o movimento browniano fracionário.

pode ser escrita como

$$p(k) \sim p_{\text{ret}}(k)r(k), \quad (1.4)$$

na qual $p_{\text{ret}}(k) \sim k^{h-2}$ é a probabilidade de retorno ao valor extremo inicial após k passos da caminhada aleatória [isto é, $B_h(t_0+k) = B_h(t_0)$] [26] e $r(k)$ representa a fração de observações no intervalo $(t_0, t_0 + k]$ visíveis a $B_h(t_0)$ [20]. Lacasa *et al.* [20] argumentaram que a função $r(k)$ deve estar associada à rugosidade da caminhada aleatória, capturada pelo seu desvio padrão. Sendo o desvio padrão do movimento browniano fracionário proporcional a t^h [24], devemos ter $r(t) \sim t^h$. Esse resultado está de acordo com a intuição de que caminhadas aleatórias mais rugosas (de incrementos antipersistentes) tendem a diminuir a visibilidade entre seus elementos. Consequentemente, a fração de elementos visíveis a um ponto extremo $B_h(t_0)$ em um intervalo de tempo k é dada por $r(k) \sim \frac{k^h}{k} = k^{h-1}$. Substituindo esses resultados na Eq. 1.4, encontramos

$$p(k) \sim k^{2h-3}. \quad (1.5)$$

Assim, comparando o expoente γ da distribuição de grau e o expoente da equação anterior, temos que

$$\gamma = 3 - 2h. \quad (1.6)$$

Usando essa relação, Lacasa *et al.* [20] concluíram que os grafos de visibilidade são uma ferramenta capaz de estimar o expoente de Hurst de séries temporais.

A Figura 1.3 ilustra a aplicação desse procedimento na estimativa do expoente de Hurst para duas trajetórias de movimento browniano fracionário. Em particular, as Figuras 1.3a e 1.3b mostram o comportamento de cada uma das séries temporais, tornando clara a relação entre a rugosidade e o caráter persistente ou antipersistente da caminhada aleatória. As Figuras 1.3c e 1.3d mostram a forma empírica da distribuição de grau $p(k)$ obtida a partir dos grafos de visibilidade associados às séries temporais comparada ao comportamento de lei de potência predito pela Eq. 1.5. Similarmente aos resultados reportados por Lacasa *et al.* [20], notamos que o comportamento das distribuições empíricas está em razoável acordo com o comportamento teórico. Entretanto, cabe ressaltar as significativas flutuações observadas nas estimativas de γ , mesmo para séries temporais relativamente longas (mais de 32 mil observações). Tais flutuações podem representar uma importante limitação na estimativa de h em séries temporais curtas.

1.1.2 Ruído, caos e grafos de visibilidade horizontal

Os grafos de visibilidade horizontal têm como principal motivação a possibilidade de se obter resultados analíticos para séries temporais de origem estocástica [8]. Em particular, Luque *et al.* [8] estimaram analiticamente o valor do coeficiente de aglomeração e do cami-

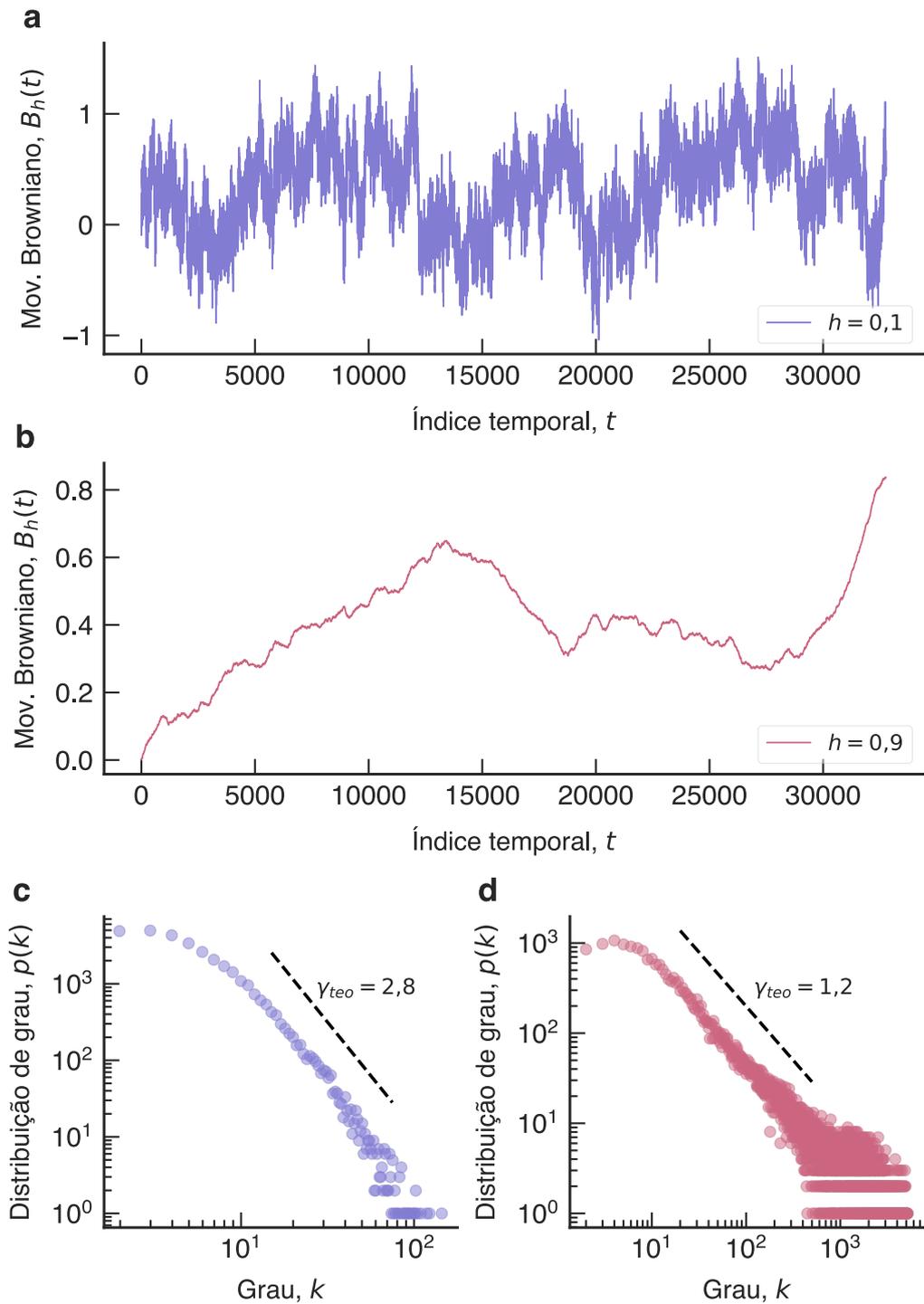


Figura 1.3: Grafo de visibilidade e movimento fracionário browniano. Séries temporais do movimento browniano fracionário com (a) $h = 0,1$ e (b) $h = 0,9$. Os painéis (c) e (d) mostram as distribuições de grau dos grafos de visibilidade construídas a partir das séries temporais em (a) e (b), respectivamente. Nesses painéis, as linhas tracejadas indicam o comportamento teórico esperado pela Eq. 1.6, próximas às quais indicamos o valor teórico do expoente da lei de potência (γ_{teo}). Ambas as realizações do processo estocástico são compostas de 2^{15} observações.

nho mais curto médio para séries temporais aleatórias. Nesse mesmo trabalho, os autores demonstraram que a distribuição de grau de um grafo de visibilidade horizontal obtido de uma sequência suficientemente longa de números aleatórios independentes e identicamente distribuídos é

$$p(k) = \frac{1}{3} \left(\frac{2}{3} \right)^{k-2}, \quad (1.7)$$

que pode ser reescrita como

$$p(k) \sim \exp(-\lambda k), \quad (1.8)$$

com $\lambda = \lambda_{\text{iid}} = \ln\left(\frac{3}{2}\right)$. Esse resultado independe da distribuição de probabilidade (desde que contínua) subjacente à sequência numérica [8].

Considerando a expressão anterior, posteriormente, Lacasa e Toral [27] investigaram a possibilidade de empregar grafos de visibilidade horizontal na tarefa de distinção entre séries temporais caóticas e séries temporais estocásticas, um problema recorrentemente encontrado no estudo de séries temporais [28, 29]. A partir de um conjunto de simulações numéricas com mapas caóticos e sequências aleatórias, Lacasa e Toral [27] concluíram que o valor do parâmetro λ da distribuição de grau (Eq. 1.8) de um grafo de visibilidade horizontal poderia ser usado para distinguir séries estocásticas de séries caóticas. Especificamente, eles verificaram numericamente a relação $\lambda_{\text{caos}} < \lambda_{\text{iid}} < \lambda_{\text{corr}}$, na qual λ_{caos} representa o valor do parâmetro λ obtido para séries temporais caóticas e λ_{corr} representa o mesmo para séries temporais estocásticas autocorrelacionadas. Desse modo, λ_{iid} representaria uma espécie de limiar de separação entre os comportamentos caótico ($\lambda_{\text{caos}} < \lambda_{\text{iid}}$) e estocástico correlacionado ($\lambda_{\text{corr}} > \lambda_{\text{iid}}$).

A Figura 1.4 ilustra como a distribuição de grau $p(k)$ (e, conseqüentemente, o valor de λ) muda ao analisarmos uma série temporal de ruído gaussiano fracionário (série de incrementos do movimento browniano fracionário) ou uma série temporal obtida da iteração do mapa logístico no regime de caos completo⁴. Além disso, mostramos também o comportamento esperado (empírico e teórico) para uma série temporal de números aleatórios não correlacionados sorteados de uma distribuição normal (ruído branco gaussiano). Podemos observar que o decaimento de $p(k)$ para o mapa logístico é consideravelmente mais lento do que para o ruído branco. Por outro lado, $p(k)$ para o ruído gaussiano fracionário vai a zero mais rapidamente que o ruído branco. Assim, os resultados da Figura 1.4 são compatíveis com a hipótese de Lacasa e Toral [27]. Entretanto, resultados mais recentes têm sugerido falhas na hipótese de que $\lambda_{\text{iid}} = \ln(3/2)$ separaria dois regimes diferentes de séries temporais. Em particular, os trabalhos de Zhang *et al.* [30] e Ravetti *et al.* [31] apontaram várias exceções à regra $\lambda_{\text{caos}} < \lambda_{\text{iid}} < \lambda_{\text{corr}}$ para diversos mapas caóticos, ruído gaussiano fracionário ($h < \frac{1}{2}$) e séries temporais de processos estocásticos autoregressivos.

⁴Mais detalhes sobre o mapa logístico são apresentados na Seção 1.3.1. Detalhes sobre o ruído gaussiano fracionário são encontrados no Apêndice B.

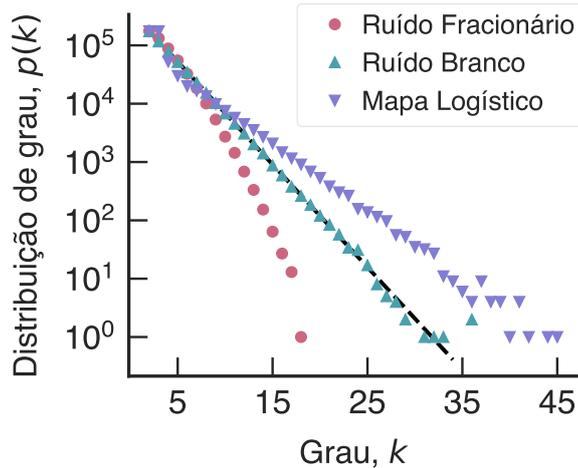


Figura 1.4: Distribuição de grau de grafos de visibilidade horizontal. Distribuição de grau para grafos de visibilidade horizontal obtidos de uma série de números aleatórios não correlacionados (ruído branco gaussiano), uma série estocástica autocorrelacionada (ruído gaussiano fracionário com $h = 0,9$) e uma série caótica (mapa logístico no regime de caos completo). Todas as séries usadas aqui são compostas de 2^{19} observações. A linha tracejada indica o comportamento da Eq. 1.8 com $\lambda = \lambda_{\text{iid}}$.

1.2 Grafos de quantil

O algoritmo de construção de um grafo de quantil é uma outra possibilidade de representação de uma série temporal como rede complexa e foi introduzido na literatura por Campanharo *et al.* [9].

Diferentemente dos grafos de visibilidade e suas variações, grafos de quantil têm o número q de vértices da rede como parâmetro do algoritmo. Mediante uma escolha conveniente de q , cada vértice representa um quantil da distribuição empírica de probabilidade dos valores da série temporal. Essa discretização da série em intervalos equiprováveis é feita por meio do ordenamento crescente dos dados de uma série temporal $\{x_t\}_{t=1,\dots,N}$, seguido do seu particionamento em q seções de mesmo tamanho. As arestas da rede capturam transições entre quantis associados a elementos consecutivos de uma série. Caso dois elementos x_i e x_{i+1} de uma série estejam nos intervalos associados aos quantis l e k , respectivamente, uma aresta direcionada ligando esses vértices existe no grafo. Vale notar que como os vértices são equivalentes entre si (devido à equiprobabilidade de ocorrência das observações associadas a cada quantil), a dinâmica da série temporal fica representada pelas arestas da rede. Por fim, o grafo de quantil é também pesado, sendo o peso de cada aresta definido pela frequência de ocorrência da transição. A Figura 1.5 ilustra o procedimento de construção de um grafo de quantil a partir de uma série temporal.

No mesmo trabalho em que propuseram os grafos de quantil, Campanharo e colaboradores [9] aplicaram essa abordagem a diversas séries temporais, como sinais periódicos simples, séries caóticas obtidas dos sistemas de Rössler e Lorenz e séries experimentais relativas a

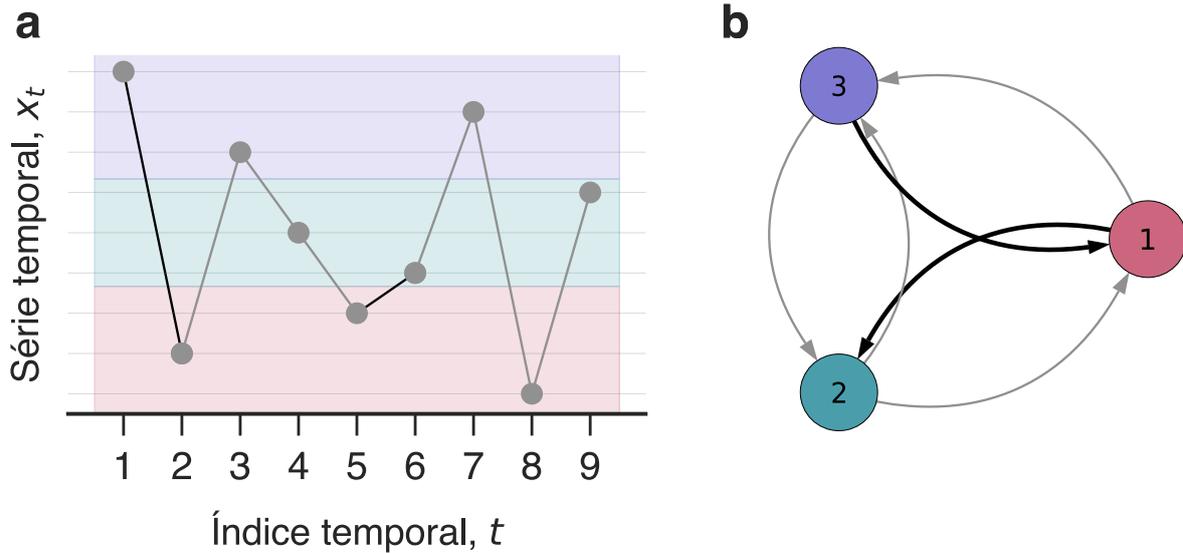


Figura 1.5: Ilustração do processo de construção de um grafo de quantil. (a) A série temporal $x_t = \{8, 1, 6, 4, 2, 3, 7, 0, 5\}$ é particionada em três quantis ($q = 3$). Por conveniência, os quantis são numerados em ordem crescente. Os segmentos mais escuros destacam duas transições: a primeira entre os quantis 3 e 1 e a segunda entre os quantis 1 e 2 (note que cada uma dessas transições ocorre uma outra vez). Os grafos de quantil permitem que vértices possam se conectar a si próprios (autoarestas), basta que dois elementos consecutivos da série temporal sejam associados ao mesmo quantil (embora isso não ocorra nesse exemplo). (b) Grafo de quantil resultante do mapeamento da série temporal x_t . As arestas mais escuras representam transições que ocorrem com maior frequência.

intervalos de tempo entre batimentos cardíacos consecutivos de pacientes saudáveis e com doenças cardíacas. Entre outros resultados, eles verificaram que grafos de quantil mapeados a partir de sinais periódicos ruidosos apresentaram uma transição entre regularidade e aleatoriedade com o aumento da amplitude do ruído, semelhante à que ocorre nas redes de “mundo pequeno” de Watts e Strogatz [32]. Mais recentemente, outros trabalhos estenderam as aplicações dessas redes de quantil a sinais de eletroencefalograma para o estudo de convulsões [33] e da doença de Alzheimer [34].

Além de possibilitarem a representação de uma série temporal como rede complexa, os grafos de quantil têm a particularidade de serem inversíveis, ou seja, é possível mapear um grafo de quantil em réplicas da série temporal original realizando uma caminhada aleatória sobre a rede correspondente à série original [9]. Desse modo, os grafos de quantil podem conectar redes complexas e séries temporais de forma dual, o que torna possível o uso de ferramentas de ambas as áreas no estudo de diversos sistemas.

1.2.1 Grafos de quantil e o expoente de Hurst

Assim como fizemos anteriormente com o grafo de visibilidade, dentre as várias aplicações dos grafos de quantil, ilustraremos seu uso na estimativa do expoente de Hurst em amostras (realizações) de movimento browniano fracionário. Em particular, vamos investigar uma abordagem baseada em transições entre quantis em diferentes escalas de tempo dentro de uma mesma série que foi recentemente desenvolvida por Campanharo e Ramos [35].

Para tanto, ao invés de construir apenas um grafo de quantil por meio do exame de elementos consecutivos de uma série temporal $\{x_t\}_{t=1,\dots,N}$, construímos uma coleção desses grafos a partir do estudo de pares de elementos espaçados em k unidades de tempo⁵, ou seja, x_i e x_{i+k} , com $k = 1, \dots, N - 1$ [33]. Assim, a estrutura de cada um desses grafos está relacionada às transições entre quantis em diferentes escalas de tempo de uma mesma série que são capturadas pela introdução do parâmetro k .

De posse de uma tal coleção de grafos, passeios aleatórios sobre os nós dessas redes permitem estudar o número de quantis que são “saltados” a cada passo do caminhante na rede. Em um grafo de quantil construído com um valor particular de k , definimos a quantidade $\delta_{l,k}(i, j) = |i - j|$ como o número de quantis saltados no l -ésimo passo de um caminhante aleatório, na qual os índices i e j representam o i -ésimo e j -ésimo quantil, respectivamente. O número médio de quantis saltados a cada passo sobre a rede é dado por⁶

$$\Delta(k) = \frac{1}{L} \sum_{l=1}^L \delta_{l,k}(i, j), \quad (1.9)$$

sendo L o número de passos da caminhada aleatória.

Campanharo e Ramos [35] observaram empiricamente que o comportamento da Eq. 1.9 para valores grandes de k é da forma $\Delta(k) \sim k^h$, possibilitando uma estimativa do parâmetro de Hurst. A Figura 1.6 mostra resultados da aplicação desse procedimento para séries temporais de movimento browniano fracionário com diferentes valores de h . Notamos que o comportamento de lei de potência é seguido de perto por $\Delta(k)$, especialmente para valores grandes de k .

1.3 Redes de recorrência

As redes de recorrência, propostas por Marwan *et al.* [12], constituem outra maneira de se representar uma sequência numérica como uma rede complexa. Essa transformação

⁵Essa generalização do procedimento de construção de um grafo de quantil foi primeiramente proposta na referência [35]. Outra exposição dessa proposta pode ser encontrada na referência [33].

⁶Uma aproximação para $\Delta(k)$ computacionalmente menos custosa pode ser encontrada na referência [33].

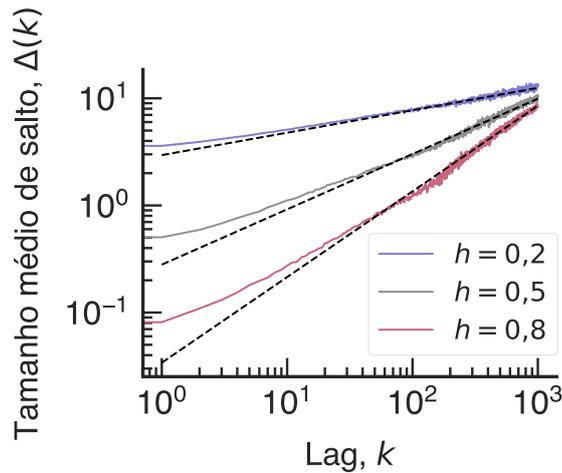


Figura 1.6: Estimativa do expoente de Hurst a partir de grafos de quantil.

Para estimar o expoente de Hurst, analisamos 50 realizações diferentes de movimento browniano fracionário para três valores diferentes de h (indicados na figura). Em seguida, a partir de cada conjunto de realizações, construímos uma coleção de grafos de quantil (para cada uma das 50 realizações de mesmo h), variando o parâmetro k desde 1 até 1000. Finalmente, investigamos caminhadas aleatórias de mesmo tamanho de cada série original para calcular $\Delta(k)$. As linhas contínuas na figura mostram o valor médio de $\Delta(k)$ para cada conjunto de realizações. As linhas tracejadas mostram ajustes dos comportamentos empíricos (via mínimos quadrados em escala log-log), os quais produzem estimativas de h bastante próximas dos valores verdadeiros (0,21 para $h = 0,2$, 0,52 para $h = 0,50$ e 0,80 para $h = 0,80$). Todas as séries estudadas eram constituídas de 2^{14} observações e usamos $q = 50$ na construção de todas as redes.

tem sua origem diretamente associada ao estudo de sistemas dinâmicos⁷, especificamente, em outra ferramenta de análise de séries temporais não lineares chamada de gráficos de recorrência [38].

Um gráfico de recorrência é uma representação pictórica de uma matriz binária e simétrica \mathbf{R} , chamada matriz de recorrência [38]. Os índices i e j , que rotulam os elementos r_{ij} em uma tal matriz, representam instantes de tempo em uma série temporal obtida de sucessivas medições/observações de um sistema. Caso os estados relativos aos instantes i e j sejam suficientemente semelhantes (próximos), dizemos que houve uma recorrência (retorno a um estado previamente ocupado) na evolução do sistema e o elemento de matriz r_{ij} é feito igual a 1. Caso os estados relativos a i e j sejam dessemelhantes, a entrada correspondente na matriz \mathbf{R} é nula. Ao representar uma matriz de recorrência por uma figura na qual os elementos de \mathbf{R} iguais a 1 são representados por células pretas e os elementos nulos por células brancas, obtemos um gráfico de recorrência [38].

⁷Um sistema dinâmico é um sistema determinístico definido por um conjunto de equações matemáticas cuja solução define a evolução do estado desse sistema [36, 37]. Quanto ao estado do sistema, esse é representado por um vetor de estado em um espaço vetorial, chamado espaço de fase, cuja dimensão depende do número de variáveis necessárias para a descrição completa do sistema. Usualmente, essas variáveis são variáveis de posição e momento generalizados.

O estado de um sistema físico (em um dado instante) usado na definição da matriz de recorrência, é representado por um ponto no espaço de fase, espaço usualmente definido pelas variáveis de posição e momento generalizados que descrevem a dinâmica do sistema. A evolução desse estado, obtida a partir da solução das equações de movimento, é representada por uma trajetória no mesmo espaço de fase. Entretanto, empiricamente e de modo geral, temos acesso a conjuntos de medições na forma de séries temporais e as regras que regem a evolução dinâmica do sistema costumam ser desconhecidas [37]. Essa aparente contradição entre teoria e prática é amenizada por métodos desenvolvidos para a análise de séries temporais não lineares [37, 39]. A ideia fundamental dessa abordagem é a possibilidade de reconstruir o espaço de fase associado a um sistema dinâmico a partir de uma série temporal $\{x_t\}_{t=1,\dots,N}$ feita uma escolha apropriada de dois parâmetros: d e τ [40]. O parâmetro d é chamado de parâmetro de *embedding* ou *embedding dimension*, enquanto τ é chamado de *lag*, *time delay* ou *embedding delay*. Esses dois parâmetros possibilitam a construção de vetores de estado da forma $\mathbf{w}_i = (x_i, x_{i+\tau}, \dots, x_{i+(d-1)\tau})$ e a similaridade entre esses vetores de estado é usada para calcular a matriz de recorrência.

Por fim, a similaridade entre dois estados \mathbf{w}_i e \mathbf{w}_j pode ser medida por uma distância. Essa medida de distância (similaridade) pode ser, por exemplo, a distância euclidiana entre os estados no espaço de fase reconstruído [12]. Nesse caso, quando a distância euclidiana entre dois vetores de estado é menor que um limite ϵ preestabelecido, ou seja, $\|\mathbf{w}_i - \mathbf{w}_j\| < \epsilon$, os estados \mathbf{w}_i e \mathbf{w}_j são suficientemente semelhantes ou recorrentes. De posse dessas definições, os elementos da matriz \mathbf{R} podem ser escritos como

$$r_{ij} = \Theta(\epsilon - \|\mathbf{w}_i - \mathbf{w}_j\|), \quad (1.10)$$

na qual $\Theta(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$ é a função degrau de Heaviside.

A partir da matriz de recorrência, obtemos uma rede de recorrência por meio de sua reinterpretação como a matriz de adjacência de um grafo simples [12]. Uma rede de recorrência é completamente definida por sua matriz de adjacência, usualmente expressa por

$$\mathbf{A} = \mathbf{R} - \mathbf{I}_N, \quad (1.11)$$

sendo \mathbf{I}_N a matriz identidade de ordem N . A subtração da matriz identidade remove as autoarestas do grafo, uma vez que todo estado é sempre recorrente consigo mesmo [12]. A conectividade da rede, por sua vez, depende do parâmetro ϵ . Valores de ϵ próximos a zero tornam a rede desconectada e valores grandes de ϵ permitem um número maior de conexões. Para escolhas apropriadas desse parâmetro, a topologia da rede fica determinada pela dinâmica de recorrências do sistema. A Figura 1.7 mostra um exemplo de uma rede de recorrência e sua matriz de adjacência.

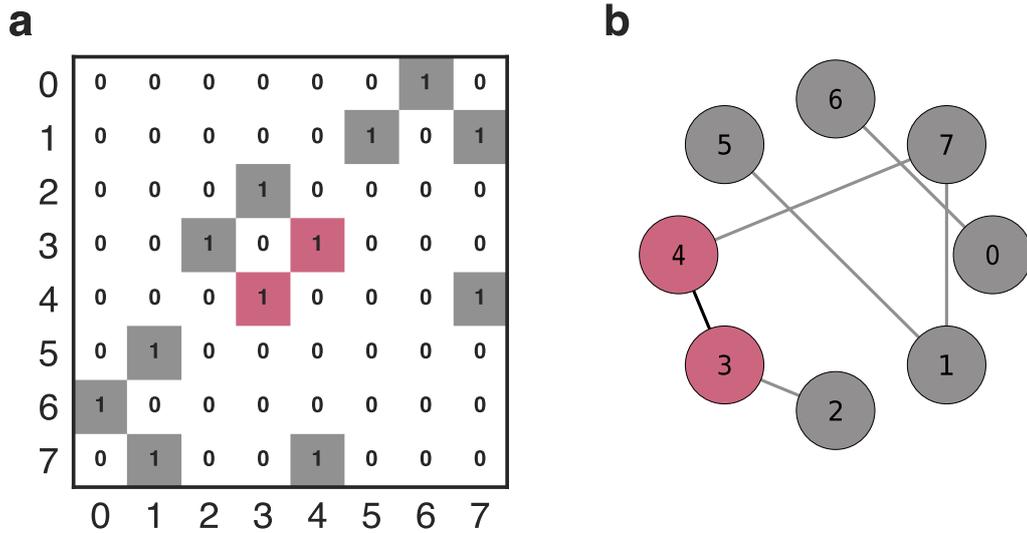


Figura 1.7: Ilustração do processo de construção de uma rede de recorrência. (a) Matriz de adjacência da rede de recorrência associada à série temporal $x_t = \{8, 1, 6, 4, 2, 3, 7, 0, 5\}$ para $\epsilon = 3$, $d = 2$ e $\tau = 1$. Os 8 vetores de estado associados a essa série são: $\{(8, 1), (1, 6), (6, 4), (4, 2), (2, 3), (3, 7), (7, 0), (0, 5)\}$. (b) Rede de recorrência resultante do mapeamento da série temporal x_t . A ligação entre os estados rotulados pelos índices 3 e 4 na matriz de adjacência é destacada.

Apesar da relação evidente entre redes de recorrência e sistemas dinâmicos, esse algoritmo tem sido amplamente empregado em investigações de séries experimentais das mais diversas áreas [5,6]. De fato, juntamente dos grafos de visibilidade, as redes de recorrência estão entre as transformações mais difundidas na literatura. Notamos ainda que as redes de recorrência são apenas um exemplo das chamadas redes de proximidade, dentre outras propostas possíveis [5]. Um algoritmo particularmente importante é devido a Zhang e Small [11] e resulta nas chamadas redes de ciclos, uma das primeiras propostas de mapeamento de séries temporais em redes complexas encontradas na literatura. Nessas redes, ciclos de séries temporais são mapeados em vértices de um grafo e ligações entre os nós são estabelecidas quando dois segmentos distintos são suficientemente semelhantes entre si. No trabalho seminal de Zhang e Small [11], o coeficiente de correlação de Pearson é usado como medida de semelhança entre ciclos de séries empíricas.

1.3.1 Redes de recorrência aplicadas ao mapa logístico

No mesmo artigo em que propõem o algoritmo de redes de recorrência, Marwan e colaboradores [12] aplicam essa transformação a séries temporais obtidas de iterações do mapa logístico [41] visando demonstrar que medidas da rede são capazes de separar diferentes regimes dinâmicos desse mapa. O mapa⁸ logístico [41] é um modelo inspirado em dinâmica

⁸Um mapa (ou equação de diferença) é um exemplo de sistema dinâmico no qual o tempo é uma variável discreta [36,37]. De um modo geral, um mapa é escrito na forma

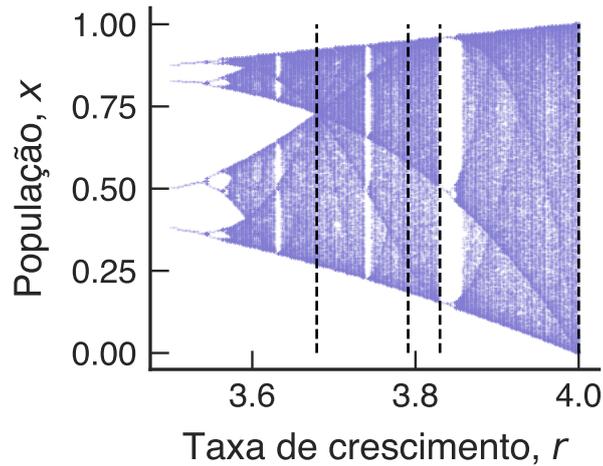


Figura 1.8: Diagrama de bifurcação do mapa logístico. Órbitas do mapa logístico para valores do parâmetro r entre 3,5 e 4,0 igualmente espaçados de 0,0015. Os valores de $r \in \{3,679; 3,791; 3,830; 4,000\}$, para os quais construímos as redes de recorrência são indicados por linhas verticais tracejadas.

populacional e definido pela equação de recorrência

$$x_{t+1} = rx_t(1 - x_t), \quad (1.12)$$

com $0 < x_t < 1$ representando uma medida hipotética de população no tempo t e r denotando o análogo de uma taxa de crescimento populacional [41, 42]. A depender do valor de r , a evolução de x_t apresenta diversos padrões dinâmicos, como oscilações periódicas e comportamento caótico.

Seguindo o trabalho de Marwan *et al.* [12], fixamos os parâmetros d e τ iguais a 1 e, para o parâmetro de similaridade ϵ , utilizamos 5% do desvio padrão de cada série temporal para $r \in \{3,679; 3,791; 3,830; 4,000\}$. Esses valores do parâmetro r foram escolhidos por corresponderem a quatro regimes diferentes do mapa, nominalmente: junção de bandas ($r = 3,679$), laminar ($r = 3,791$), periódico ($r = 3,830$) e caos completo ($r = 4,000$). A Figura 1.8 mostra um diagrama de bifurcação do mapa logístico destacando esses valores de r . Para cada um dos quatro valores de r , iteramos o mapa 10000 vezes (com condição inicial aleatória) e descartamos as primeiras 1000 iterações a fim de evitar comportamentos transientes. Em seguida, utilizamos esses dados para construir quatro redes de recorrência (para cada uma das séries referentes aos quatro valores de r) usando os parâmetros mencionados anteriormente. De posse dessas redes, estimamos o comprimento médio do caminho mais curto \mathcal{L} , o coeficiente de aglomeração \mathcal{C} e a densidade ρ da rede⁹, conforme mostra a

$$\mathbf{x}_{t+1} = \mathbf{F}(\mathbf{x}_t)$$

com $\mathbf{x} = (x_1, x_2, \dots, x_n)$ e \mathbf{F} um campo vetorial de mesma dimensão [36, 37].

⁹A densidade ρ de uma rede é a razão entre o número de arestas que ocorrem na rede e o número de máximo de arestas que poderia ocorrer. O coeficiente de aglomeração \mathcal{C} e o valor médio do caminho mais

Tabela 1.1.

Um aspecto interessante dessas métricas é que cada uma delas tem clara interpretação no contexto das recorrências. O comprimento médio do caminho mais curto entre dois vértices da rede representa a distância entre dois estados no espaço de fase reconstruído. Isso implica em restrições diretas a essa medida, uma vez que estados mais distantes no espaço de fase também devem ser distantes entre si na rede. Por conta dessa característica, essas redes de recorrência tendem a não apresentar a propriedade de “mundo pequeno” [32]. O coeficiente de aglomeração, por sua vez, representa a probabilidade de que os vizinhos de um determinado estado sejam próximos entre si no espaço de fase. Por fim, a densidade da rede corresponde a uma medida global de recorrência do sistema. Vale mencionar que os resultados reproduzidos na Tabela 1.1 estão em ótimo acordo com o trabalho original de Marwan *et al.* [12] e, combinados, permitem a distinção entre os quatro regimes investigados do mapa logístico.

Regime	r	\mathcal{L}	\mathcal{C}	ρ
junção de bandas	3,679	22,8	0,83	0,051
laminar	3,791	23,3	0,78	0,040
periódico	3,830	1,00	1,00	0,333
caótico	4,000	23,6	0,82	0,045

Tabela 1.1: Propriedades das redes de recorrência para séries temporais do mapa logístico em diferentes regimes. Aqui, \mathcal{L} representa o comprimento médio do caminho mais curto entre pares de vértices da rede, \mathcal{C} é o coeficiente de aglomeração e ρ a densidade da rede. O parâmetro r determina o regime do mapa logístico.

1.4 Redes Ordinais

As redes ordinais, recentemente propostas por Small [10], constituem o último exemplo de rede complexa mapeada de série temporal apresentado nesta dissertação. Essas redes podem ser entendidas como uma extensão teórica do formalismo associado a uma medida de complexidade de séries temporais introduzida por Bandt e Pompe [43] chamada entropia de permutação. Sendo assim, antes de descrevermos o algoritmo da transformação, começamos por fazer uma breve explanação dessa medida de complexidade.

Dada uma série temporal $\{x_t\}_{t=1,\dots,N}$, podemos construir $n = N - d + 1$ diferentes partições superpostas de comprimento d . Conseqüentemente, essas partições contém $d - 1$ observações em comum. O parâmetro d , que define o tamanho das partições, é chamado de *embedding dimension* e sua escolha é limitada pelas condições $d! \ll N$ e $d \geq 2$ [43]. Escolhido d , cada partição de x_t é então denotada por um vetor $\mathbf{w}_s = (x_s, x_{s+1}, \dots, x_{s+d-1})$ com $s = 1, 2, \dots, n$. A cada elemento em \mathbf{w}_s associamos um número do vetor de ordem $(0, 1, \dots, d - 1)$. Esse número

curto \mathcal{L} são definidos no Apêndice A, no qual apresentamos mais detalhes acerca de redes complexas.

é relativo à posição de cada elemento em \mathbf{w}_s de modo que à observação x_s é associado o número 0, a x_{s+1} o número 1 e assim sucessivamente até que todos os elementos em \mathbf{w}_s estejam associados a um índice do vetor de ordem. Em seguida, procuramos pela permutação $\pi_s = (r_0, r_1, \dots, r_{d-1})$ do vetor de ordem $(0, 1, \dots, d-1)$ que ordena crescentemente os elementos da série que compõem a partição. Por fim, a permutação π_s é associada à partição \mathbf{w}_s . A fim de ilustrar esse procedimento, consideramos uma partição hipotética $\mathbf{w} = (3, 7, 5)$ obtida de uma série temporal $\{x_t\}$. Como $d = 3$, o vetor de ordem correspondente é $(0, 1, 2)$. Para que \mathbf{w} tenha seus elementos ordenados crescentemente, os números 7 e 5 têm de trocar de posição dentro da partição, ou seja, devem ser permutados. Aplicando essa permutação ao segundo e terceiro elementos do vetor de ordem, associamos a permutação $\pi = (0, 2, 1)$ a essa partição visto que essa é a permutação que ordena \mathbf{w} .

A ideia de ordenar os elementos de uma partição implica na possibilidade de estabelecer desigualdades claras entre os mesmos, o que não ocorre caso dois ou mais elementos de uma partição sejam iguais entre si. Uma possível solução, adotada ao longo deste trabalho, consiste em alterar minimamente a organização crescente do vetor de ordem $(0, 1, \dots, d-1)$ não permutando elementos repetidos dentro de uma partição [44, 45]. Para esclarecer essa escolha, supomos uma partição $\mathbf{w} = (4, 3, 4)$. O vetor de ordem é novamente $(0, 1, 2)$. À primeira observação do número 4 associamos o número de ordem 0. À sua segunda observação cabe o índice 2. Uma simples permutação do número 3 com o número 4 (primeiro elemento da partição) ordena \mathbf{w} . Logo, $\pi = (1, 0, 2)$ é uma permutação que ordena essa partição. Note, entretanto, que uma segunda permutação entre os números 4 não altera o ordenamento crescente da janela e $(1, 2, 0)$ seria outra permutação legítima para ordenar \mathbf{w} . Para evitar qualquer ambiguidade na definição de π , escolhemos manter a ordem de ocorrência das observações, ou seja, definimos $\pi = (1, 0, 2)$ como a permutação associada a \mathbf{w} .

Repetido o processo anterior para cada um dos \mathbf{w}_s , obtemos uma sequência de permutações, também chamada de sequência simbólica ou sequência de padrões ordinais $\{\pi_s\}_{s=1, \dots, n}$. De posse dessa sequência simbólica, podemos calcular a probabilidade de ocorrência da permutação π_i como

$$p_i(\pi_i) = \frac{\text{número de ocorrências de } \pi_i \text{ em } \{\pi_s\}_{s=1, \dots, n}}{n}. \quad (1.13)$$

Conseqüentemente, a partir da aplicação da Eq. 1.13, obtemos a distribuição de probabilidade empírica dessas permutações, representada por $P = \{p_i(\pi_i)\}_{i=1, \dots, d!}$. É importante notar que o parâmetro d (*embedding dimension*) limita a $d!$ o número máximo de diferentes padrões ordinais (permutações) contidos em $\{\pi_s\}_{s=1, \dots, n}$.

O cálculo da entropia de Shannon a partir da distribuição de probabilidade das permutações

tações é o que chamamos de entropia de permutação, representada por

$$H = - \sum_{i=1}^{d!} p_i \ln p_i. \quad (1.14)$$

A entropia de permutação tem seu mínimo ($H = 0$) para uma série monotônica, na qual apenas um padrão ordinal aparece na sequência simbólica. Seu máximo ($H = \ln d!$) ocorre quando todas as permutações relacionadas à série temporal são equiprováveis, o que é observado para sequências de números aleatórios [43]. Essencialmente, a entropia de permutação pode ser pensada como uma medida da variedade de padrões ordinais encontrados em uma série [46]. Assim, Bandt e Pompe introduziram a entropia de permutação como uma medida de complexidade em séries temporais.

Subseqüentes desenvolvimentos e aplicações da entropia de permutação levaram à adição de outro parâmetro no processo de simbolização. Esse parâmetro é conhecido como *embedding delay*¹⁰ e usualmente denotado pela letra grega τ , com $\tau \geq 1$ [44, 45]. Enquanto na proposta original da entropia de permutação elementos consecutivos de uma série temporal formavam as partições nas quais uma série era dividida, esse novo parâmetro possibilita a formação de partições constituídas de observações não consecutivas, distantes mais de uma unidade de tempo entre si, possibilitando o estudo de padrões ordinais em diversas escalas de tempo¹¹.

A introdução do parâmetro τ faz conveniente revisitar o processo de particionamento de uma série. Considerando uma série temporal $\{x_t\}_{t=1, \dots, N}$, podemos dividi-la em $n' = N - (d - 1)\tau$ partições. Cada uma dessas partições é representada por um vetor $\mathbf{w}'_s = (x_s, x_{s+\tau}, \dots, x_{s+(d-1)\tau})$ com $s = 1, 2, \dots, n'$. Assim como fizemos anteriormente, um vetor de ordem $(0, 1, \dots, d - 1)$ é associado a cada janela e procuramos pela permutação π_s que ordena as observações da série que constituem cada janela. Uma distribuição de probabilidade pode então ser obtida da sequência simbólica resultante assim como a entropia de Shannon.

É a partir desse contexto de uma série temporal codificada em uma sequência simbólica que surgem as redes ordinais. Ao invés de procurarmos pela distribuição de probabilidade dos padrões ordinais e calcularmos a entropia, construímos uma rede das permutações. Nessa rede, cada vértice representa uma das diferentes permutações presentes em $\{\pi_s\}_{s=1, \dots, n}$ e as arestas ligam pares de permutações que aparecem consecutivamente na sequência ordinal. Desse modo, as arestas são direcionadas da permutação que ocorre primeiro à permutação

¹⁰Os formalismos da entropia de permutação e das matrizes de recorrência compartilham dos mesmos parâmetros porque ambas ferramentas estão diretamente ligadas a ideias subjacentes à análise de séries temporais não lineares [39]. Uma completa junção desses dois formalismos acontece através dos chamados gráficos de recorrência de ordem [47]. Nessas figuras, semelhantes aos gráficos de recorrência, os estados relativos às entradas da matriz de recorrência são definidos pelas permutações do processo de simbolização de Bandt e Pompe. As recorrências são avaliadas a partir da igualdade dessas permutações [47].

¹¹O parâmetro k introduzido na construção dos grafos de quantil para a análise do parâmetro de Hurst pode ser pensado como o análogo de τ naquele formalismo.

ção seguinte, capturando a sucessão temporal dos padrões ordinais. Além de direcionadas, essas ligações são também pesadas e pares de permutações cujas transições ocorrem mais frequentemente apresentam ligações mais fortes, sendo o peso da ligação igual a probabilidade de observação dessa transição na sequência simbólica. Formalmente, a probabilidade de transição ($i \rightarrow j$) entre dois vértices i e j da rede é

$$p_{ij} = \frac{\text{número de vezes que } \pi_i \text{ é seguida de } \pi_j \text{ em } \{\pi_s\}_{s=1,\dots,n}}{n-1}. \quad (1.15)$$

Conhecidas as probabilidades de transição entre todos os pares de permutações, podemos determinar a matriz de adjacência e construir a rede ordinal.

Entretanto, o procedimento anterior de construção de uma rede ordinal, continuação natural das ideias de Bandt e Pompe, não condiz com a proposta original desses grafos. No algoritmo original, recentemente proposto por Small [10], observações adjacentes de uma série são particionadas em janelas não sobrepostas, diferentemente do particionamento original da entropia de permutação. Após a proposta de Small, vários outros algoritmos de formação de redes ordinais, ligeiramente diferentes, têm surgido na literatura. Em uma primeira variação, usando já de partições sobrepostas, os vértices não representam apenas permutações, mas contém informação referente ao símbolo de permutação e um símbolo associado à amplitude dos números da partição, semelhantemente aos grafos de quantil [48]. Em outra generalização, cujo procedimento de construção usa dos parâmetros d e τ , a estrutura dessas redes foi investigada para valores de τ maiores que 1 [15, 17]. Em todos os algoritmos, entretanto, as arestas capturaram a sucessão temporal dos vértices (permutações), caracterizando as redes ordinais como um tipo de rede de transição. A Figura 1.9 mostra um exemplo de rede ordinal construída com janelas superpostas, $d = 3$ e $\tau = 1$.

Assim como acontece com redes de recorrência, redes ordinais têm sido empregadas principalmente no estudo de sistemas dinâmicos [10, 14, 15, 48], mas aplicações a sistemas biológicos, mais especificamente, séries de intervalos entre batimentos cardíacos [17, 18], também são encontradas na literatura. A possibilidade de inversão dessa transformação também foi investigada [14]. De modo similar aos grafos de quantil, o procedimento de inversão passa pela realização de uma caminhada aleatória sobre os vértices da rede. Por meio de escolhas apropriadas dos parâmetros de *embedding* é possível construir séries temporais “substitutas” cuja dinâmica é próxima à dinâmica da série original [14].

1.4.1 Redes ordinais e o sistema de Rössler

Conforme já mencionamos, as redes ordinais têm sido especialmente aplicadas ao estudo de séries temporais não lineares. Dentre essas aplicações, destacamos seu uso na análise de diferentes regimes do sistema de Rössler [49], um conjunto de equações diferenciais acopladas

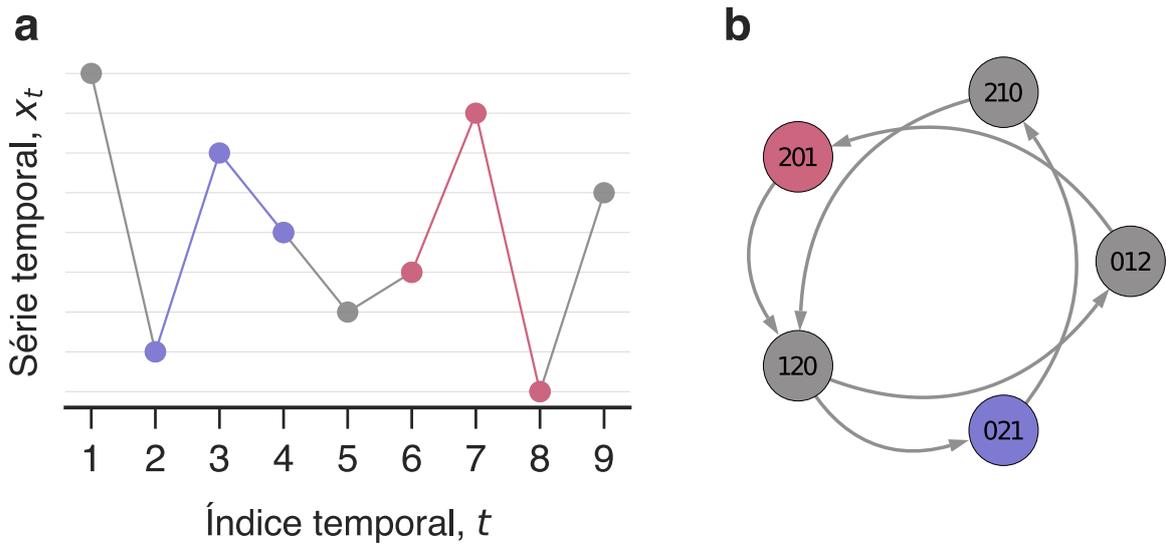


Figura 1.9: Ilustração do procedimento de construção de rede ordinal. (a) A série $x_t = \{8, 1, 6, 4, 2, 3, 7, 0, 5\}$ é mapeada em uma rede ordinal escolhendo $d = 3$ e $\tau = 1$. Note que escolhemos sobrepor as partições nas quais a série é dividida. Duas permutações diferentes, $(2, 0, 1)$ e $(0, 2, 1)$, correspondentes às janelas $\mathbf{w}_2 = (1, 6, 4)$ e $\mathbf{w}_6 = (3, 7, 6)$ são destacadas. Cinco das seis permutações possíveis (dadas a escolha de d) ocorrem nessa curta sequência. (b) Rede ordinal mapeada de x_t . Os vértices correspondentes às permutações $\pi_2 = (2, 0, 1)$ e $\pi_6 = (0, 2, 1)$ associadas às janelas \mathbf{w}_2 e \mathbf{w}_6 são destacados.

definido por

$$\begin{aligned} \frac{dx}{dt} &= -y - z \\ \frac{dy}{dt} &= x + ay \\ \frac{dz}{dt} &= b + z(x - c) \end{aligned} \quad (1.16)$$

Esse sistema de equações foi introduzido com o objetivo de definir um modelo simples de dinâmica não linear no qual o tempo é uma variável contínua¹². A complexidade observada ao longo da evolução temporal desse sistema é consequência da relação não linear do produto zx presente em sua última equação [49].

No mesmo artigo em que introduz a transformação de redes de transição de partições ordinais, Small [10] constrói redes ordinais a partir de séries temporais da variável x para

¹²Sistemas dinâmicos nos quais o tempo é uma variável contínua são descritos por conjuntos de equações diferenciais da forma

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{F}(\mathbf{x}(t)).$$

Esses sistemas dinâmicos são também chamados de fluxos [36, 37]. Na equação acima, $\mathbf{x} = (x_1, x_2, \dots, x_n)$ e \mathbf{F} é um campo vetorial de mesma dimensão. No caso do sistema de Rössler, em que o campo vetorial não depende explicitamente do tempo, o sistema é ainda chamado de autônomo [36, 37].

diferentes regimes do sistema de Rössler. Para tanto, são fixados os parâmetros $b = 2$ e $c = 4$ e as mudanças de regime são delegadas ao parâmetro a . Apesar da análise realizada se restringir somente ao estudo da forma não direcionada e não pesada dessas redes, Small argumenta que uma simples inspeção visual da topologia das redes seria capaz de demonstrar diferenças entre regimes periódicos e caóticos [10].

Em trabalho posterior, McCullough e colaboradores [15], construindo a rede direcionada com janelas sobrepostas e usando os parâmetros d e τ , variaram o parâmetro $a \in \{0,37; 0,3705; 0,371; \dots; 0,43\}$ novamente com $b = 2$ e $c = 4$ e notaram que, mediante certas escolhas de d e τ , séries periódicas da variável x do sistema de Rössler geravam redes de estrutura circular. Séries caóticas dessa mesma variável transformavam-se em grafos de aparência tubular, em alusão ao atrator do sistema. De modo a investigar sistematicamente se simples métricas de uma rede ordinal seriam capazes de capturar alterações na dinâmica do sistema, os autores escolhem usar $d = 8$ e $\tau = 9$ de modo que a estrutura da rede contenha o máximo de informação possível e ao mesmo tempo seja uma conveniente simplificação da dinâmica do sistema. Variando o parâmetro a no intervalo mencionado, um diagrama de bifurcação é produzido (novamente da variável x) e o expoente de Lyapunov¹³ λ estimado para cada uma das séries. O comportamento do valor médio do grau de saída $\langle k \rangle$ e da variância do grau de saída σ_k da rede ordinal correspondente são mostrados na Figura 1.10. Como podemos ver, essas medidas simples e globais da rede parecem se comportar semelhantemente ao expoente de Lyapunov e, portanto, capturam alterações na dinâmica do sistema subjacente.

¹³O expoente de Lyapunov [42] é uma medida da divergência entre trajetórias de um sistema dinâmico com condições iniciais ligeiramente distintas. Usando a definição de caos dada por Strogatz [42], de que “caos é comportamento aperiódico de longo prazo em um sistema determinístico que exibe *sensível dependência às condições iniciais*”, o expoente de Lyapunov é uma quantificação da sensibilidade do sistema a alterações nas condições iniciais.

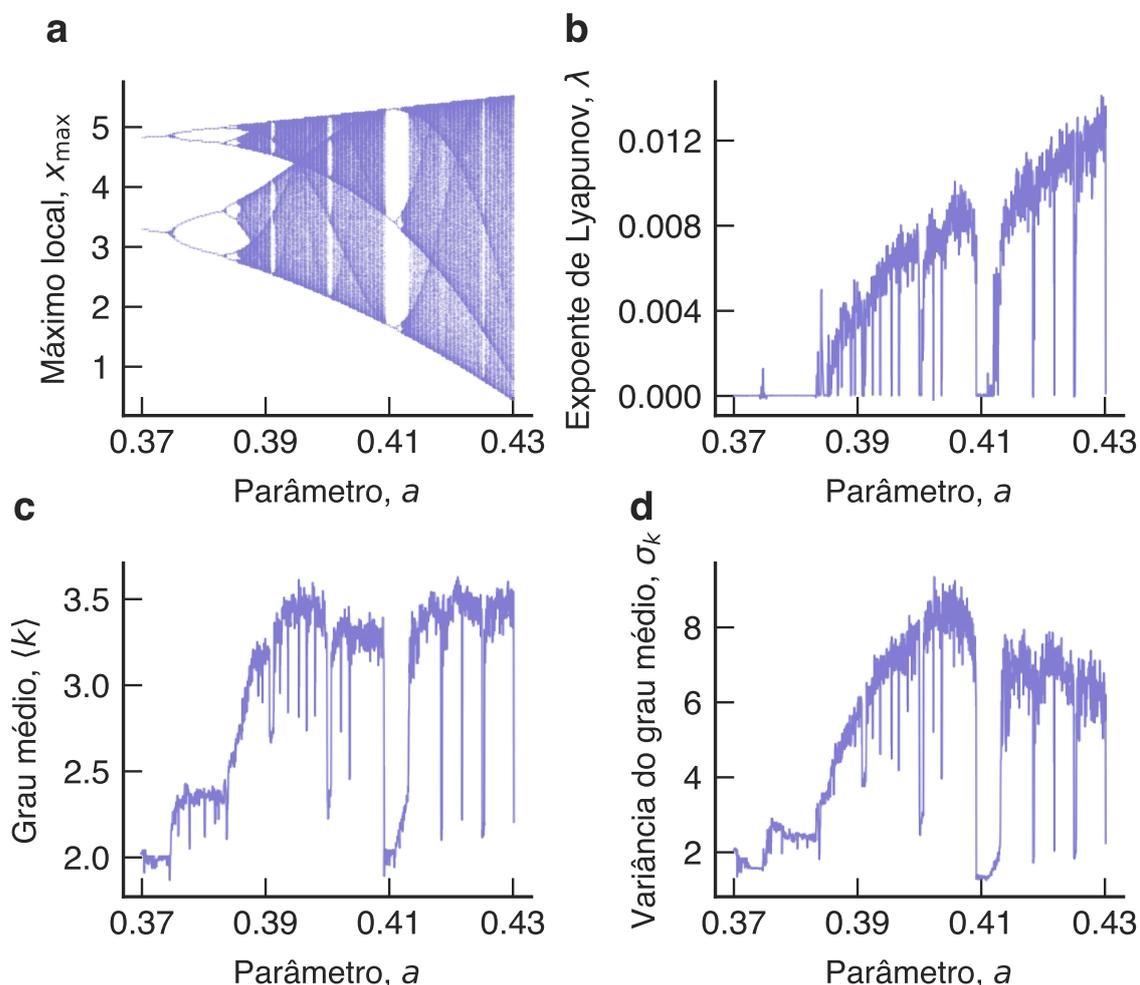


Figura 1.10: Alterações na dinâmica do sistema de Rössler. (a) Diagrama de bifurcação para a variável x do sistema de Rössler. Esse sistema de equações foi resolvido para 1201 valores igualmente espaçados do parâmetro a entre 0,37 e 0,43 usando o método de Runge-Kutta de quarta ordem [50]. Para cada valor diferente desse parâmetro, um conjunto de condições iniciais aleatórias no intervalo $(0, 1)$ é gerado para as variáveis x , y e z . O passo de integração escolhido foi 0,02 e foram realizadas 200000 iterações. Metade desses pontos foi descartada para evitar transientes. Ainda, amostramos a cada 10 pontos das soluções do sistema de equações a fim de reduzir as séries a 10000 dados. Os resultados mostrados foram obtidos dessas últimas séries. (b) Expoente de Lyapunov em função da variação do parâmetro a . (c) Grau médio de saída de redes ordinais construídas usando $d = 8$ e $\tau = 9$ em função do parâmetro a . (d) Variância do grau de saída para as mesmas redes ordinais. O resultados mostrados nessa figura são uma reprodução parcial de resultados de McCullough e colaboradores [15].

Propriedades básicas e aplicações de redes ordinais a processos estocásticos

Neste segundo capítulo, apresentamos novos resultados para as redes ordinais [19]. Primeiramente, tornamos explícitos alguns vínculos inerentes a essas redes que implicam em diretas restrições de conectividade. Em seguida, identificamos aspectos da topologia das redes de sinais simples bem como a estrutura de redes ordinais mapeadas de séries temporais de natureza estocástica. Por fim, aplicamos as redes ordinais à análise de autocorrelações em séries temporais estocásticas teóricas e empíricas.

2.1 Estrutura e propriedades de redes ordinais

Assim que uma série temporal é mapeada em uma rede complexa, as métricas da rede passam a ser as métricas relevantes para o estudo da série. Sendo assim, é importante sabermos como o processo de mapeamento de uma série influencia na estrutura da rede complexa. Características topológicas importantes das redes ordinais são naturalmente herdadas e vinculadas às sequências simbólicas que dão origem a essas redes. Por isso, passamos a explicitar algumas restrições importantes impostas às redes ordinais [10, 15, 19, 48].

A conectividade de um vértice é uma característica singular nas redes ordinais. Essas redes apresentam restrições quanto ao número máximo de arestas que saem ou chegam a um vértice. Curiosamente, essas restrições não são aparentes quando $d = 2$, pois nesse caso a rede contém todas as conexões direcionadas possíveis entre os nós da rede (Figura 2.1 e Figura 2.2b). Entretanto, apesar de termos seis permutações possíveis em uma rede ordinal quando $d = 3$, um máximo de três arestas saem ou chegam a cada vértice. Essa limitação

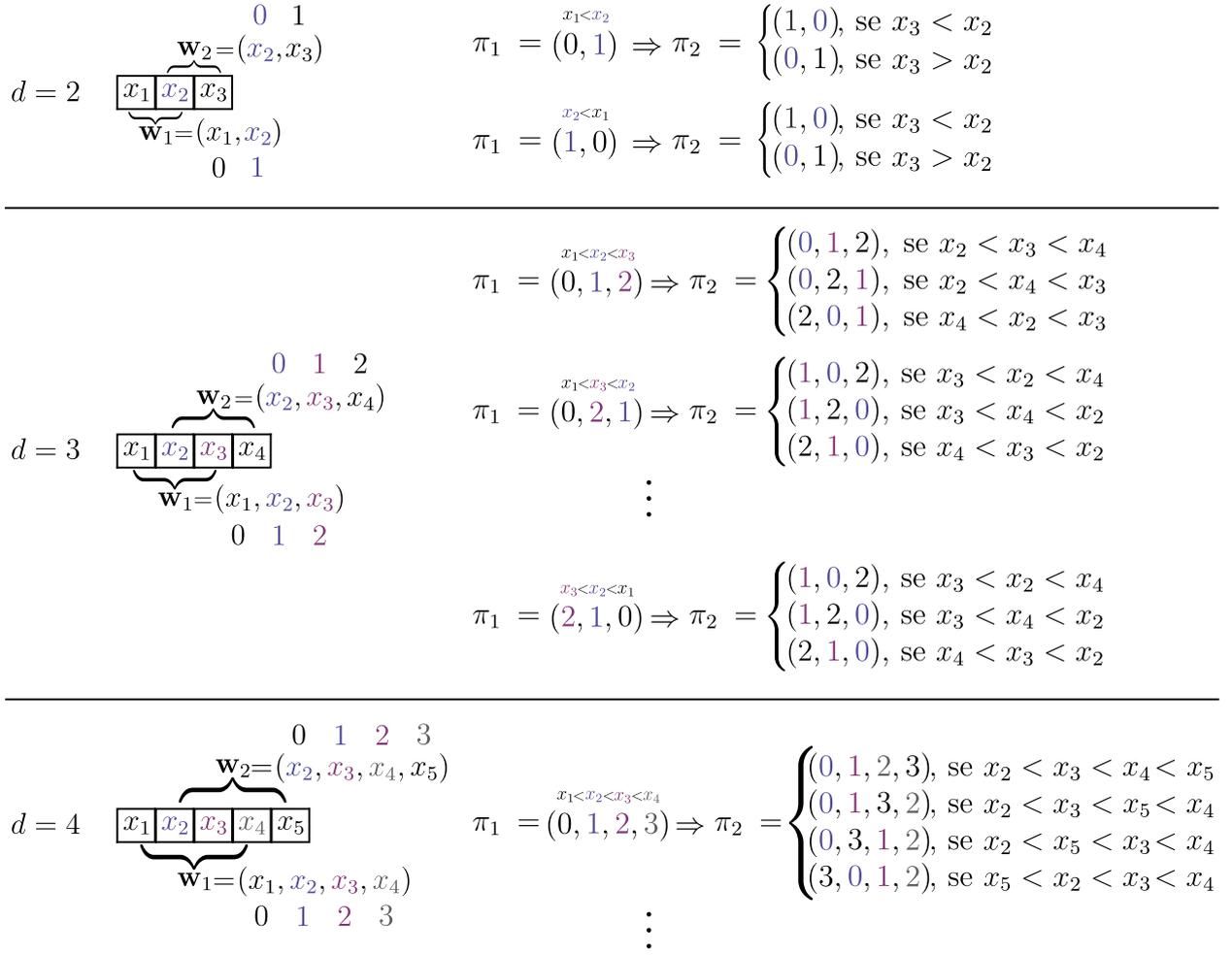


Figura 2.1: Transições permitidas em redes ordinais. Alguns exemplos de transições possíveis em redes ordinais de *embedding dimension* $d = 2, 3$ e 4 são mostradas na figura. Observações x_1, x_2, \dots de uma série temporal que integram duas partições consecutivas da série são destacadas em cor. Além disso, os números de ordem associados a essas observações (em cada uma das janelas) são correspondentemente coloridos. Note que o símbolo ordinal associado a cada observação da série decresce com o avanço da janela móvel.

de conectividade deve-se ao fato de que a permutação π_s que ordena a partição \mathbf{w}_s de uma série temporal pode ser sucedida por somente três padrões ordinais associados à partição \mathbf{w}_{s+1} . Esse vínculo imposto às permutações é consequência do ordenamento dos elementos na partição \mathbf{w}_s , dado que tal ordenamento é parcialmente levado adiante para a próxima partição \mathbf{w}_{s+1} pelos dois elementos que ambas compartilham (Figura 2.1). Quanto menor a *embedding dimension* d , menos observações são compartilhadas entre partições diferentes e o passado de uma série temporal se torna mais rapidamente irrelevante na determinação de futuras permutações. Apesar de termos discutido apenas o caso $d = 3$, essas ideias de sucessão permanecem válidas para qualquer escolha de d , de modo que todos os nós em uma rede de permutações têm graus de entrada e saída limitados a valores inteiros entre 0 e d .

Desse modo, uma rede ordinal tem o número total de arestas limitado pelo produto entre o número máximo de arestas que um vértice pode apresentar e o número máximo de vértices que a rede pode ter, ou seja, limitado a $d \times d!$.

Outra consequência dos vínculos impostos às transições entre permutações está relacionada à ocorrência de autoarestas nessas redes. Analisando todas as transições possíveis (respeitando os vínculos de sucessão), concluímos que autoarestas podem existir para apenas dois vértices de uma rede ordinal, independentemente da *embedding dimension* d . Esses dois nós são aqueles associados a permutações ascendentes ou descendentes, como as permutações $(0, 1, 2)$ e $(2, 1, 0)$ no caso em que $d = 3$ e $(0, 1, 2, 3)$ e $(3, 2, 1, 0)$ para $d = 4$.

Como diferentes maneiras de se construir redes ordinais podem ser encontradas na literatura, os vínculos anteriores não se aplicam diretamente a essas generalizações. Por exemplo, uma dada permutação π_s pode ser seguida de qualquer permutação no caso de partições consecutivas que não se sobrepõem. Isso ocorre pois essas partições não contém elementos comuns entre si [10, 16]. No caso em que o *embedding delay* τ entre as observações da série é maior que 1, restrições similares às discutidas anteriormente emergem para transições de ordem mais alta [15]. Nesta dissertação, entretanto, vamos nos concentrar em redes ordinais que representam um desdobramento natural do processo de simbolização de Bandt e Pompe [43], com janelas sobrepostas, *embedding dimension* d e *embedding delay* $\tau = 1$.

2.1.1 Topologia de redes ordinais de séries temporais simples

Discutidas as restrições gerais de conectividade das redes ordinais, passamos a apresentar resultados acerca da topologia dessas redes para séries temporais simples.

Talvez o primeiro e mais simples exemplo a ser considerado seja o de uma sequência monotônica (crescente ou decrescente). Nesse caso, independentemente da *embedding dimension* d , a rede ordinal é formada por apenas um nó (representando uma permutação crescente ou decrescente) e uma autoaresta (indicando que essa permutação é sempre seguida de si mesma). Portanto, as propriedades de uma rede ordinal mapeada de uma série monotônica são todas triviais.

Séries periódicas são outro exemplo de sinal simples mas que já conduzem a resultados mais interessantes. Redes ordinais de séries periódicas formam estruturas cíclicas, com o arranjo dos nós e arestas evocando a própria natureza da série, como mostrado na Figura 2.2a. Se o número de observações dentro de um período de uma série temporal é T e escolhermos $d = T$, a rede ordinal resultante é um grafo cíclico de T vértices (número de diferentes permutações), dado que a série temporal (e, conseqüentemente, os padrões ordinais) se repete a cada T observações. Além disso, a estrutura cíclica da rede não muda se aumentarmos a *embedding dimension* de modo que $d > T$, apenas as permutações associadas aos nós da rede são alteradas. Em razão desse comportamento, redes ordinais de sinais periódicos apresentam distribuição de grau uniforme e diâmetro que cresce linearmente com d (Figuras 2.2c e 2.2d).

Podemos mostrar que o caminho mais curto pesado médio para um vértice qualquer da rede é igual a $\frac{1}{d} \sum_{i=1}^{d-1} \left(\frac{i}{d}\right)$, resultado que também independe da escolha do nó devido à simetria do grafo. Sendo assim, o valor médio do caminho mais curto pesado para toda a rede é também

$$\langle l \rangle_{\text{per}} = \frac{1}{d} \sum_{i=1}^{d-1} \left(\frac{i}{d}\right) = \frac{d-1}{2d}, \quad (2.1)$$

com a soma representando a distância pesada de um nó qualquer da rede em relação a todos os outros e o fator $1/d$ associado à média sobre os nós da rede. Notamos que essa expressão é válida caso todas as arestas do grafo apresentem o mesmo peso, o que acontece quando a série temporal periódica é longa ou composta de um número inteiro de períodos.

2.1.2 Redes ordinais aleatórias

Sequências de números aleatórios independentes e identicamente distribuídos são outro tipo comum de série temporal que passamos a investigar. Chamaremos de redes ordinais aleatórias as redes ordinais mapeadas de ruído branco.

Diferentemente de sinais monotônicos ou periódicos, esperamos que as redes ordinais aleatórias sejam tão conectadas quanto possível, visto que os vértices dessas redes representam permutações que ocorrem equiprovavelmente em sinais aleatórios suficientemente longos [43]. Essa intuição inicial é confirmada e as redes ordinais aleatórias são formadas por $d!$ nós com d ligações saindo e d ligações chegando a cada um deles (Figura 2.2b). Logo, as redes ordinais aleatórias apresentam máxima conectividade, guardadas as restrições de sucessão intrínsecas ao processo de simbolização de Bandt e Pompe. Tal fato faz com que o grau médio e o diâmetro das redes ordinais aleatórias aumente linearmente com d , como mostrado nas Figuras 2.2c e 2.2d.

Outro detalhe interessante das redes ordinais aleatórias é que os pesos das arestas são diferentes, isto é, nem todas as transições ocorrem com a mesma probabilidade (diferentemente do que acontece com as permutações). De fato, para cada vértice da rede, uma das d arestas que sai do vértice é duas vezes mais pesada (arestas pretas na Figura 2.2b) que cada uma das $d - 1$ arestas restantes. Empiricamente, verificamos que esse resultado é válido para longas sequências de números aleatórios sorteados de distribuições de probabilidade de suporte contínuo. Para esclarecer o mecanismo que gera essa diferença entre as probabilidades de transição, escolhemos $d = 3$ e supomos uma partição $\mathbf{w}_1 = (x_1, x_2, x_3)$ que tem como permutação correspondente $\pi_1 = (0, 1, 2)$ (isto é, $x_1 < x_2 < x_3$). A partição seguinte, $\mathbf{w}_2 = (x_2, x_3, x_4)$, tem x_4 como único elemento não compartilhado com \mathbf{w}_1 . Necessariamente, x_4 tem de satisfazer uma das seguintes condições: *i*) $x_4 < x_1 < x_2 < x_3$; *ii*) $x_1 < x_4 < x_2 < x_3$; *iii*) $x_1 < x_2 < x_4 < x_3$; *iv*) $x_1 < x_2 < x_3 < x_4$. As condições *i* e *ii* levam à permutação $\pi_2 = (2, 0, 1)$, enquanto *iii* resulta em $\pi_2 = (0, 2, 1)$ e *iv*

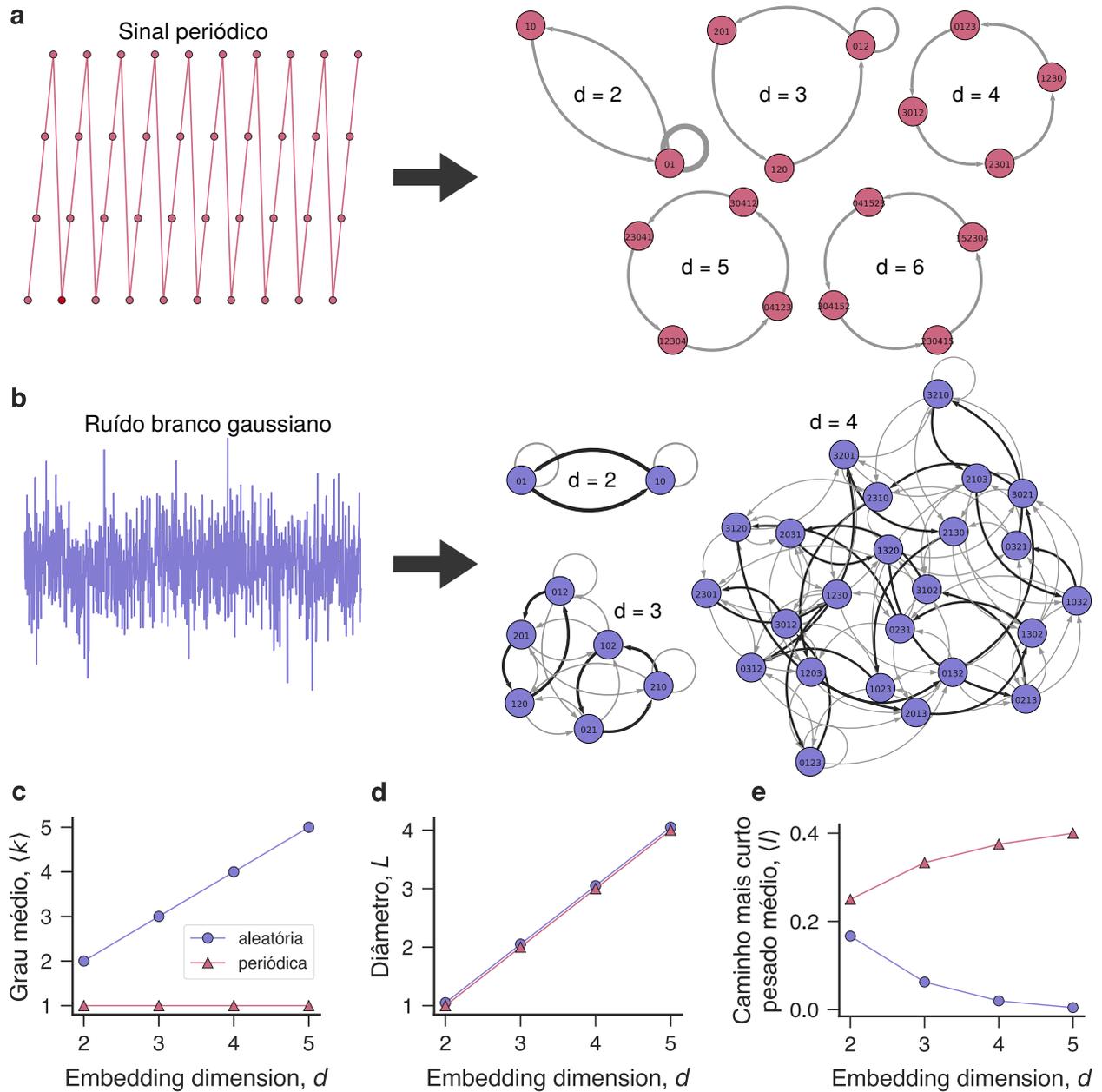


Figura 2.2: Redes ordinais de séries temporais periódicas e aleatórias. (a) Ilustração do mapeamento de um sinal periódico em redes ordinais de diferentes *embedding dimensions* d (indicadas na figura). (b) Mapeamento de ruído branco gaussiano em redes ordinais de diferentes *embedding dimensions* d . As arestas pretas indicam as transições que ocorrem com maior probabilidade. (c) Grau médio $\langle k \rangle$, (d) diâmetro L e (e) caminho mais curto pesado médio $\langle l \rangle$ em função da *embedding dimension* d para redes ordinais de sinais aleatórios (círculos) e periódicos (triângulos). Os sinais periódicos têm período T e *embedding dimension* d iguais. Os resultados para essas três propriedades são válidos para séries temporais suficientemente longas, quando temos uma estimativa confiável de todas as transições entre as permutações.

em $\pi_2 = (0, 1, 2)$. Assim, existem duas maneiras de encontrarmos $\pi_2 = (2, 0, 1)$, o que torna a transição $(0, 1, 2) \rightarrow (2, 0, 1)$ duas vezes mais provável que $(0, 1, 2) \rightarrow (0, 2, 1)$ ou $(0, 1, 2) \rightarrow (0, 1, 2)$. Vale notar que se sortearmos um número grande de ternas (x_1, x_2, x_3) , os valores médios de x_1, x_2 e x_3 convergem para os quartis da distribuição de probabilidade da série temporal, tornando as quatro condições anteriores equiprováveis. Essas mesmas ideias podem ser aplicadas para transições de primeira ordem qualquer que seja a *embedding dimension* d .

Uma regra simples e geral para determinar a permutação π_{s+1}^* que é mais provável de seguir π_s é procurar pela permutação na qual o símbolo igual a ' $d-1$ ' encontra-se na posição ocupada pelo símbolo '0' em π_s . Por exemplo, se $d = 4$ e $\pi_s = (3, 2, 1, 0)$, temos que $\pi_{s+1}^* = (2, 1, 0, 3)$ é duas vezes mais provável de suceder $(3, 2, 1, 0)$ que qualquer uma das outras permutações possíveis $\{(3, 2, 1, 0), (2, 1, 3, 0), (2, 3, 1, 0)\}$, visto que o símbolo '3' em π_{s+1}^* está na mesma posição ocupada pelo símbolo '0' em π_s . Esse resultado, juntamente das restrições de sucessão entre permutações, permite construir as matrizes de adjacência teóricas para as redes ordinais aleatórias de *embedding dimension* d arbitrária.

Para calcularmos as entradas dessas matrizes de adjacência teóricas, começamos com uma matriz quadrada nula de ordem $d!$. Estabelecemos, então, conexões direcionadas entre todos os pares de nós π_i e π_j para os quais a transição $\pi_i \rightarrow \pi_j$ é possível. Os elementos de matriz correspondentes a essas arestas são feitos iguais a 1. Em seguida, atualizamos os pesos das ligações mais prováveis de 1 para 2. Finalmente, os elementos p_{ij} da matriz de adjacência são divididos pelo fator $(d+1)!$, que representa o peso total das arestas da rede $[(d-1) \times d! + 2d!]$. Como exemplo, caso $d = 2$, a matriz de adjacência teórica de uma rede ordinal aleatória é dada por

$$\mathbf{A}_{d=2} = \begin{pmatrix} \frac{1}{(2+1)!} & \frac{2}{(2+1)!} \\ \frac{2}{(2+1)!} & \frac{1}{(2+1)!} \end{pmatrix} = \begin{pmatrix} 1/6 & 2/6 \\ 2/6 & 1/6 \end{pmatrix}, \quad (2.2)$$

de modo que os pesos das ligações são $p_{01 \rightarrow 01} = 1/6$, $p_{01 \rightarrow 10} = 2/6$, $p_{10 \rightarrow 01} = 2/6$ e $p_{10 \rightarrow 10} = 1/6$.

Matrizes de adjacência construídas similarmente a $\mathbf{A}_{d=2}$ permitem calcular teoricamente o resultado de qualquer métrica de uma rede ordinal aleatória de *embedding dimension* d arbitrária. A Figura 2.2e mostra, por exemplo, que o valor médio do caminho mais curto pesado em função de d (estimado a partir dessas redes teóricas) tende a zero com o crescimento de d . Isso acontece porque os pesos das arestas decrescem muito rapidamente com o aumento de d .

2.2 Sinal, ruído e redes ordinais aleatórias

As matrizes de adjacência teóricas das redes ordinais aleatórias tornam possível a aplicação dessas redes a situações que envolvam a presença (ou ausência) de ruído em uma série temporal.

Como primeiro exemplo de aplicação, investigamos a possibilidade de detecção de comportamento regular em séries aleatórias curtas. Para ilustrar essa possibilidade, investigamos se ruído branco gaussiano parcialmente ordenado pode ser distinguido de um ruído “puro” estimando a distância de edição¹ δ [51] de suas respectivas redes ordinais. Para uma rede complexa direcionada e pesada, essa medida simples corresponde à quantidade de peso das arestas que deve ser realocado para que as duas redes se tornem iguais. Assim sendo, geramos um *ensemble* de séries de ruído branco nas quais uma fração η de elementos consecutivos são ordenados crescentemente. Em seguida, variando a fração de dados ordenados η , mapeamos essas séries temporais em redes ordinais e estimamos os valores médios da distância de edição δ dessas redes em relação às redes aleatórias teóricas. A Figura 2.3 mostra o comportamento do valor médio de δ em função de η , bem como o intervalo de 95% de confiança para $d \in \{2, 3, 4\}$ e séries compostas de $N \in \{1000, 5000, 10000\}$ elementos. Como esperado, observamos que os valores de δ aumentam com η em todos os casos.

Também estimamos o valor médio da distância de edição entre redes ordinais empíricas de ruído branco (1000, 5000 e 10000 observações) e redes aleatórias teóricas para construir bandas de confiança do comportamento de séries aleatórias finitas. Valores de δ fora dessa banda representam uma indicação confiável de que uma série temporal empírica apresenta desvios em relação à completa aleatoriedade. Além disso, essas regiões de confiança permitem identificar frações limiares de ordenamento η^* a partir das quais a distância de edição é capaz de detectar precisamente essa anomalia. A Figura 2.3 mostra que essa abordagem é capaz de detectar regularidade em séries temporais aleatórias de 1000 elementos quando $\eta^* \approx 6,8\%$ para $d = 2$, $\eta^* \approx 8,0$ quando $d = 3$ e $\eta^* \approx 9,2$ para $d = 4$. A Figura 2.3 ainda indica que os valores de η^* diminuem com o aumento do comprimento N da série temporal e que $d = 2$ apresenta os menores valores de η^* , independentemente de N .

Para confirmar esses resultados, estimamos valores médios de η^* sobre 10 realizações do procedimento de detecção em função de N . A Figura 2.4 mostra que η^* decai exponencialmente com N para $d \in \{2, 3, 4\}$. Para séries temporais longas ($N > 10^5$), todos esses valores de d são igualmente eficientes em detectar a fração mínima de ordenamento usada em nossos experimentos numéricos ($\eta_{\min} = 1\%$). Por outro lado, $d = 2$ exibe os menores valores de η^*

¹A distância de edição δ entre dois grafos (G e G') é definida por

$$\delta(G, G') = \sum_{i,j=1}^n |a_{ij} - a'_{ij}|,$$

na qual a_{ij} denota os elementos da matriz de adjacência do grafo G e a'_{ij} representa o mesmo para o grafo G' .

para séries mais curtas. As poucas possibilidades de transição quando $d = 2$ tornam o desequilíbrio que é inserido na série (número desproporcional de padrões ordinais ascendentes) mais facilmente detectável. Entretanto, a proximidade entre as bandas de desvio padrão de $d = 2$ e $d = 3$ indica que a diferença entre esses casos é pequena. Além disso, essa situação deve mudar caso o padrão anômalo inserido na série seja mais complexo e redes ordinais com $d > 2$ podem ter um melhor desempenho na detecção de anomalias mais complexas. Esse exemplo ilustra que sempre existirá um *trade-off* entre a *embedding dimension* e o tamanho da série temporal.

Como segundo exemplo de aplicação das redes ordinais aleatórias, passamos a investigar uma medida de entropia da rede chamada de entropia global dos vértices² [17]. Essa medida captura a diversidade de sucessão entre as permutações da rede e tem sido usada como um indicador de determinismo em séries temporais [18]. Entretanto, antes de introduzir a entropia global dos vértices, começamos por definir uma entropia local, calculada ao nível do nó. Essa entropia local do vértice pode ser escrita para um vértice i da rede como

$$h_i = - \sum_{j \in \mathcal{O}_i} p'_{ij} \log p'_{ij}, \quad (2.3)$$

com $p'_{ij} = p_{ij} / \sum_{k \in \mathcal{O}_i} p_{ik}$ denotando a probabilidade renormalizada da transição $i \rightarrow j$ e \mathcal{O}_i representando o conjunto dos nós j ligados a i pelas arestas que saem desse nó (vizinhança de saída do nó i). A probabilidade renormalizada p'_{ij} implica em $0 < h_i < \log d$, com o limite inferior refletindo uma situação de completa certeza, quando temos apenas uma transição $i \rightarrow j$ possível. Por outro lado, o limite superior corresponde à situação de maior incerteza, quando temos um conjunto de transições ($i \rightarrow j, i \rightarrow k, \dots$) equiprováveis³.

Estabelecida a entropia local, podemos definir a entropia global dos vértices como

$$H_{\text{GN}} = \sum_{i=1}^d h_i p'_i, \quad (2.4)$$

na qual $p'_i = \sum_{j \in \mathcal{I}_i} p_{ji}$ representa a soma dos pesos das arestas que chegam ao nó i através de sua vizinhança de entrada \mathcal{I}_i . Logo, vemos que a entropia global dos vértices representa uma média do determinismo local das conexões da rede ponderada pelas probabilidades de ocorrência dos vértices. Notamos ainda que $p'_i \sim p_i(\pi_i)$ para séries temporais longas, ou seja, a soma dos pesos das arestas que chegam a um vértice representa a probabilidade de observação de sua permutação correspondente. Desse modo, a rede ordinal contém em sua estrutura a informação necessária para o cálculo da entropia de permutação.

²Originalmente proposta por Unakafov e Keller [46], essa medida é também chamada de entropia condicional de padrões ordinais.

³Note que o limite superior da entropia local do nó é determinado pela *embedding dimension* d , visto que o número máximo de arestas que podem sair do vértice é igual a d .

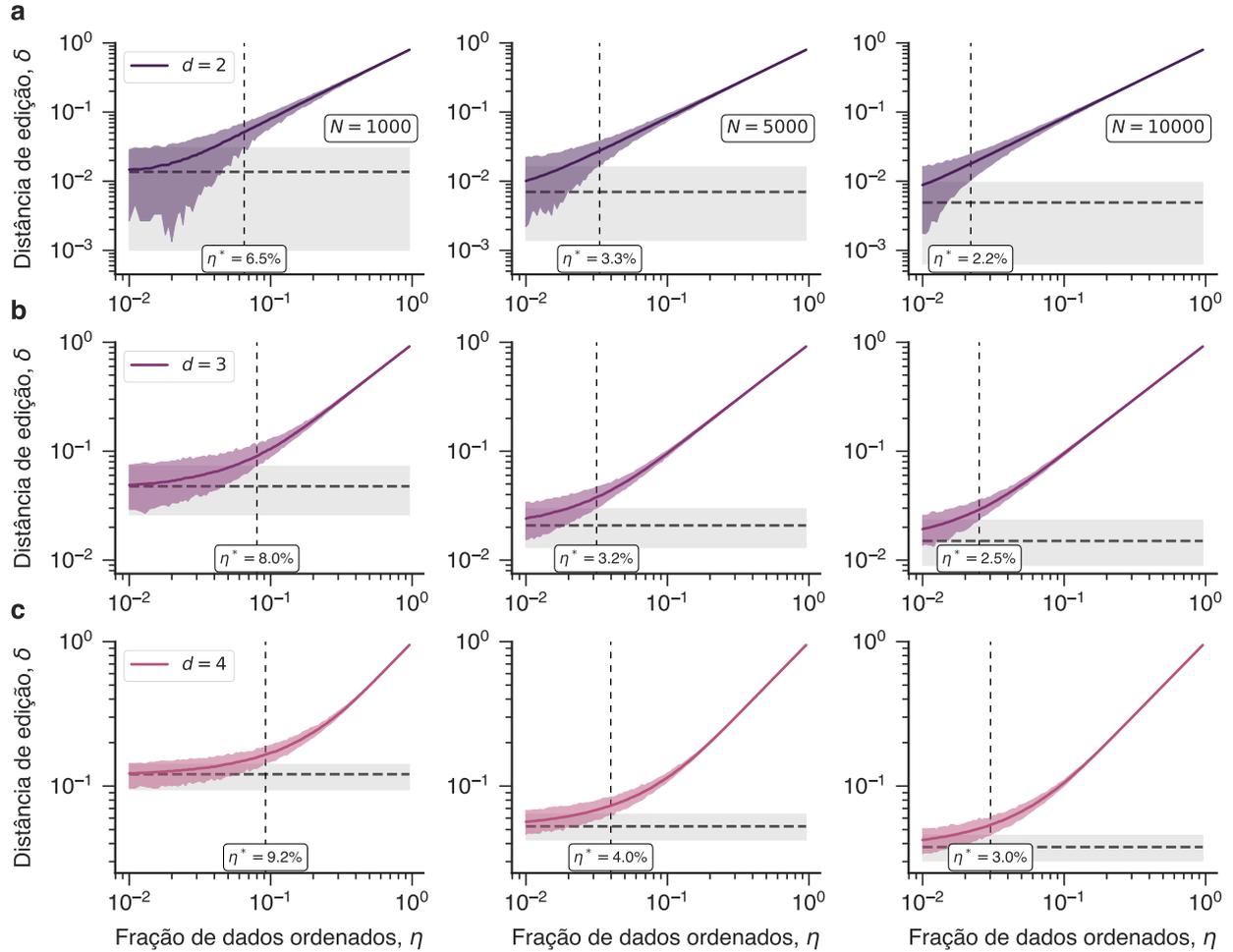


Figura 2.3: Detectando comportamento não aleatório com redes ordinais aleatórias. Os painéis (a)-(c) mostram os valores médios da distância de edição δ (linhas em cor) entre redes ordinais de ruído branco parcialmente ordenado (η é a fração de elementos ordenados) e as redes aleatórias calculadas teoricamente para as *embedding dimensions* $d = 2, 3$, e 4 , respectivamente. Em cada um desses painéis, os resultados da primeira coluna são obtidos de ensembles de 1000 séries temporais de $N = 1000$ elementos cada, enquanto as segunda e terceira colunas mostram os resultados obtidos para ensembles de 1000 séries compostas de $N = 5000$ e 10000 elementos, respectivamente. As regiões sombreadas em cor representam bandas de confiança de 95% para δ das séries parcialmente ordenadas e as regiões sombreadas em cinza denotam bandas de confiança de 95% para δ calculadas de séries aleatórias finitas. As linhas verticais tracejadas indicam a fração limiar de elementos ordenados η^* a partir da qual é possível distinguir uma série aleatória parcialmente ordenada de uma série aleatória de mesmo comprimento.

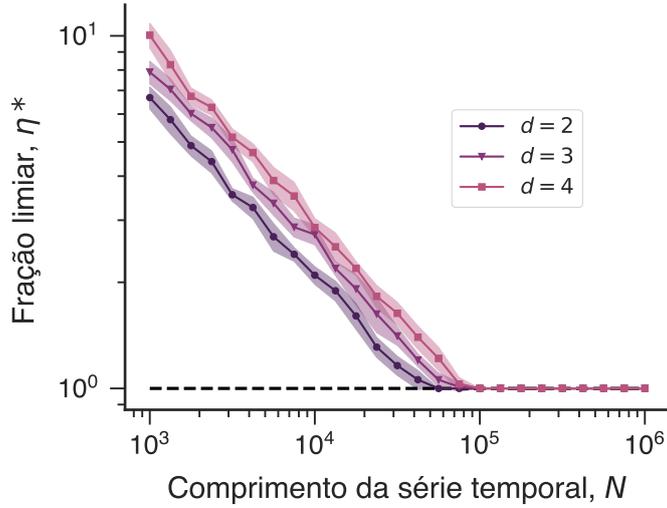


Figura 2.4: Limiar de detecção de séries aleatórias parcialmente ordenadas. Fração limiar η^* em função do comprimento N da série temporal para $d \in \{2, 3, 4\}$. Os pontos representam valores médios de η^* estimados a partir de 10 realizações e as regiões sombreadas mostram a banda de um desvio padrão. A linha tracejada horizontal indica o mínimo de ordenamento imposto em nossos experimentos numéricos (1%).

Definida a entropia da rede ordinal, seu valor teórico pode ser calculado para uma rede ordinal aleatória de *embedding dimension* d arbitrária. Primeiramente, a entropia local de um vértice i da rede é

$$\begin{aligned}
 h_i^{(\text{rand})} &= -\tilde{p}'_{ij} \log \tilde{p}'_{ij} - \sum_{j=1}^{d-1} p'_{ij} \log p'_{ij} \\
 &= -\frac{2}{d+1} \log \left(\frac{2}{d+1} \right) - \left(\frac{d-1}{d+1} \right) \log \left(\frac{1}{d+1} \right),
 \end{aligned} \tag{2.5}$$

na qual $p'_{ij} = 1/(d+1)$ representa a probabilidade renormalizada de transição do nó i para o nó j (para as transições $i \rightarrow j$ menos prováveis) e $\tilde{p}'_{ij} = 2p'_{ij}$ representa a mesma medida para a transição mais provável restante. Como os vértices em uma rede aleatória teórica têm todos o mesmo número de ligações e distribuição de pesos, podemos substituir o resultado anterior, juntamente com $p'_i = p_i^{(\text{rand})} = 1/d!$ (equiprobabilidade de ocorrência das permutações), na definição da entropia global dos vértices (Eq. 2.4) a fim de encontrarmos a entropia global H_{GN} para sequências longas de números aleatórios não correlacionados, que fica

$$\begin{aligned}
 H_{\text{GN}}^{(\text{rand})} &= \sum_{i=1}^{d!} \left\{ \frac{2}{(d+1)d!} \log \left(\frac{d+1}{2} \right) + \left(\frac{d-1}{(d+1)d!} \right) \log (d+1) \right\} \\
 &= \log(d+1) - (\log 4)/(d+1).
 \end{aligned} \tag{2.6}$$

É importante observar que o valor numérico de $H_{\text{GN}}^{(\text{rand})}$ é sempre menor que aquele obtido

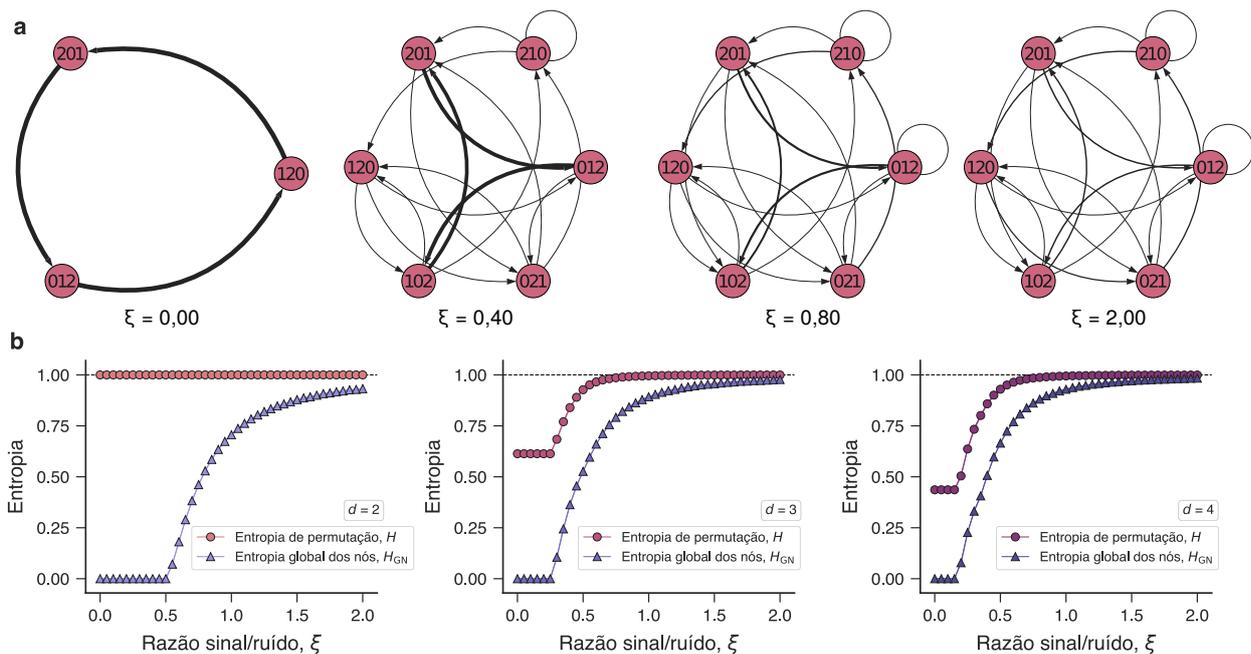


Figura 2.5: Robustez da entropia de permutação e da entropia global dos vértices à adição de ruído. (a) Visualizações de redes ordinais mapeadas de sinais periódicos ruidosos para $d = 3$ e diferentes valores da razão sinal/ruído ξ (mostrados abaixo das redes). (b) Valores médios da entropia de permutação H (círculos) e entropia global dos nós H_{GN} (triângulos) obtidos a partir de um *ensemble* de 100 realizações de sinais periódicos ruidosos em função de ξ para $d = 2, 3$, e 4 . Cada série temporal tem 10^4 elementos e o período T é feito igual à *embedding dimension* ($T = d$). A entropia de permutação é normalizada pelo seu valor máximo e a entropia global dos vértices pelo valor calculado para a rede ordinal aleatória (Eq. 2.6). As pequenas regiões sombreadas correspondem a uma banda de um desvio padrão.

no caso em que todas as transições são equiprováveis ($H_{GN}^{(equi)} = \log d$) e somente quando $d \rightarrow \infty$ temos $H_{GN}^{(rand)} \rightarrow H_{GN}^{(equi)}$. Assim, o valor de $H_{GN}^{(rand)}$ não representa a situação de mais alta entropia para um dado valor de d , visto que podem existir séries temporais para as quais $H_{GN} > H_{GN}^{(rand)}$. De posse do resultado anterior, podemos analisar como redes ordinais de sinais periódicos se transformam em redes ordinais aleatórias com a adição de ruído aos sinais. Para tanto, geramos sinais oscilatórios tipo dente de serra (Figura 2.2a) com 10000 elementos e período T . Como exemplos dessas séries, temos $x_t = \{0, 1, 0, 1, \dots\}$ ($T = 2$) e $x_t = \{0, 1/2, 1, 0, 1/2, 1, \dots\}$ ($T = 3$). Em seguida, somamos um ruído uniformemente distribuído no intervalo $[-\xi, \xi]$ aos elementos dessas séries temporais, com o parâmetro ξ controlando a razão sinal/ruído. A Figura 2.5 mostra exemplos de redes ordinais obtidas desses sinais periódicos para $\xi \in \{0; 0,4; 0,8; 2,0\}$ e $d = T = 3$. Como esperado, a estrutura circular da rede ordinal do sinal periódico ($\xi = 0$) aproxima-se de uma rede ordinal aleatória com o aumento de ξ . Notamos que essa aproximação inicia-se com a emergência de todas as permutações (o que acontece para pequenos valores de ξ), seguida do enfraquecimento das três ligações da rede cíclica inicial e o fortalecimento das outras ligações.

Essas séries periódicas ruidosas permitem avaliar a robustez à adição de ruído da entropia global dos vértices H_{GN} (Eq. 2.4) em comparação com a entropia de permutação H (Eq. 1.14). Com esse propósito, geramos 100 sinais do tipo dente de serra com $N = 10000$ dados para cada $\xi \in \{0; 0,05; 0,1; \dots; 2,0\}$. Em seguida, estimamos os valores médios de H_{GN} e H para todos os valores de ξ . Por fim, normalizamos essas quantidades dividindo os valores de H_{GN} por H_{GN}^{rand} e H por $\log d!$. A Figura 2.5b mostra a média dos valores normalizados de H_{GN} e H em função de ξ para $d = 2, 3$ e 4 . Notamos que os valores de H_{GN} são mais robustos do que os de H em relação à adição de ruído e, portanto, mais eficientes em distinguir séries com diferentes valores de ξ . Essa característica é mais evidente para $d = T = 2$ pois, nesse caso, $x_t = \{0; 1; 0; 1; \dots\}$ e os padrões ordinais são equiprováveis para todos os valores de ξ , de modo que os valores normalizados de H são sempre iguais a um. Contrariamente, a entropia global dos vértices é zero para $\xi < 0,5$ e começa a aumentar para valores de ξ maiores que $0,5$. Para valores maiores de d , observamos que a entropia de permutação H tende a 1 muito mais rapidamente do que H_{GN} . Por exemplo, ao contrário dos valores de H_{GN} , os valores de H são incapazes de distinguir essas séries periódicas ruidosas se $\xi > 1$.

2.3 Redes ordinais e correlações em séries temporais

Após analisarmos sequências de números aleatórios não correlacionados, passamos a explorar amostras de processos estocásticos autocorrelacionados. Com esse objetivo, estudamos séries de ruído gaussiano fracionário e movimento browniano fracionário [24]. Como mencionamos anteriormente, o movimento browniano fracionário é um processo estocástico autossimilar caracterizado pelo parâmetro ou expoente de Hurst $h \in (0, 1)$. Seus incrementos constituem um processo estocástico estacionário de distribuição normal de média zero e variância unitária, usualmente chamado de ruído gaussiano fracionário⁴.

A Figura 2.6 traz exemplos de realizações de ambos os processos estocásticos para diferentes valores de h , com amostras geradas usando o procedimento de Hosking⁵ [52]. A Figura 2.6 também apresenta visualizações das redes ordinais correspondentes a cada realização dos processos. Observamos que todas as permutações possíveis estão presentes nessas redes e que cada vértice apresenta todas as conexões permitidas. Desse modo, essas redes são todas idênticas se não considerarmos o peso das ligações. Entretanto, uma simples inspeção visual nos padrões dos pesos das arestas já informa acerca das particularidades de cada série temporal. Para o ruído gaussiano fracionário, observamos uma distribuição desigual dos pesos para valores pequenos do expoente de Hurst h , enquanto o movimento browniano fracionário apresenta uma distribuição mais uniforme para a mesma faixa de valores de h . No caso do movimento browniano fracionário, notamos que o peso da autoaresta associada à

⁴No Apêndice B apresentamos diversas propriedades desse processo estocástico.

⁵No Apêndice B discutimos os detalhes da implementação desse método.

permutação ‘210’ cresce consideravelmente para valores grandes de h , refletindo a tendência descendente dessas realizações do processo estocástico.

Uma possibilidade de quantificar essa aparente desigualdade nos pesos das arestas é calcular o coeficiente de Gini [53]. Esse coeficiente representa uma medida de dispersão em distribuições de probabilidade. Valores do índice de Gini próximos de zero indicam que os pesos estão igualmente distribuídos, enquanto valores próximos de um indicam uma desigualdade aguda na distribuição dos pesos. Estimamos o coeficiente de Gini de um *ensemble* de redes ordinais associadas a séries temporais de ruído gaussiano fracionário e movimento browniano fracionário com diferentes valores do parâmetro de Hurst (100 realizações para cada $h \in \{0,1; 0,2; \dots; 0,9\}$). Os valores médios do coeficiente de Gini são mostrados na Figura 2.6c e confirmam a análise visual anterior. Para o ruído gaussiano fracionário, observamos que o índice de Gini decresce quando h aumenta de 0,1 até aproximadamente 0,5. Para $h > 0,5$, o valor numérico desse coeficiente fica próximo do resultado esperado para uma rede aleatória teórica. Contrariamente, o coeficiente de Gini aumenta sistematicamente com h para o movimento browniano fracionário.

Avaliamos os valores médios da entropia global dos vértices H_{GN} (normalizada pela Eq. 2.5) em comparação com a entropia de permutação H (normalizada por $\log d!$). A Figura 2.6d mostra os resultados encontrados para o ruído gaussiano fracionário. Notamos que a dependência de H com h é mais côncava do que a dependência de H_{GN} com h e que os valores de H_{GN} apresentam um intervalo de variação muito maior do que os valores de H . Esse comportamento é similar ao reportado para sinais periódicos ruidosos e dá suporte à hipótese de que a entropia global dos vértices possui um poder discriminador maior do que a entropia de permutação usual. Ainda, calculamos a média do caminho mais curto pesado $\langle l \rangle$ para o movimento browniano fracionário em função do expoente de Hurst. A Figura 2.6e mostra que os valores de $\langle l \rangle$ decrescem monotonicamente com h para $d = 2$ e $d = 3$, um comportamento que também é observado para o ruído gaussiano fracionário.

O comportamento monotônico e bem definido de $\langle l \rangle$ em relação a h sugere que podemos usar a média do caminho mais curto pesado de uma rede ordinal para prever o expoente de Hurst da série original. Para testar sistematicamente essa possibilidade, propomos uma tarefa de regressão na qual o expoente de Hurst do movimento browniano fracionário é predito mapeando uma série em uma rede ordinal e usando $\langle l \rangle$ como única variável preditiva. Para realizar essa tarefa de regressão, escolhemos um algoritmo de k -primeiros vizinhos⁶ [54]. Esse algoritmo de aprendizagem estatística é especialmente adequado para situações em que temos uma relação não linear de baixa dimensionalidade (Figura 2.6e) [54].

Com o objetivo de prever h , geramos um *ensemble* contendo 100 amostras de movimento browniano fracionário (1024 passos) para cada $h \in \{0,1; 0,12; 0,14; \dots; 0,9\}$ e utilizamos 75%

⁶No Apêndice C expomos aspectos gerais sobre aprendizagem estatística e o algoritmo de primeiros vizinhos.

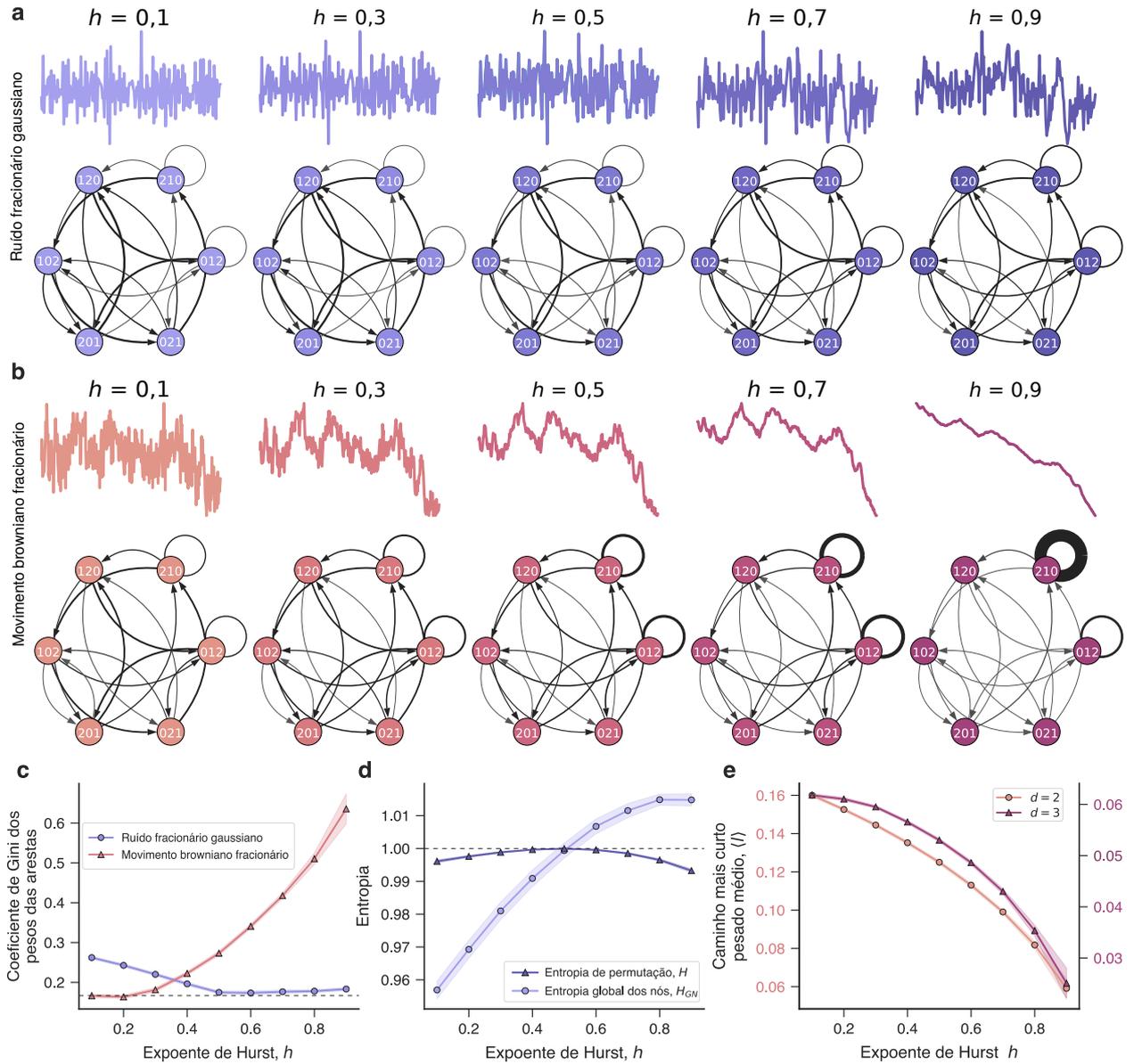


Figura 2.6: Redes ordinais de ruído gaussiano fracionário e movimento browniano fracionário. (a) Ilustração do mapeamento de amostras de ruído gaussiano fracionário com diferentes expoentes de Hurst h em rede ordinais. (b) Ilustração do mapeamento de amostras de movimento browniano fracionário com diferentes expoentes de Hurst h em redes ordinais. Nesses dois painéis, a espessura das ligações é proporcional ao peso das arestas. (c) Dependência do coeficiente de Gini associado à distribuição de peso das arestas com o expoente de Hurst h para o ruído gaussiano fracionário (círculos) e o movimento browniano fracionário (triângulos). A linha tracejada nesse painel representa o valor do coeficiente de Gini para redes ordinais aleatórias. (d) Entropia de permutação H (triângulos) e entropia global dos vértices H_{GN} (círculos) com $d = 3$ em função do expoente de Hurst h para o ruído gaussiano fracionário. A entropia de permutação é normalizada por $\log d!$ e a entropia global dos vértices pelo valor das redes ordinais aleatórias (Eq. 2.6). (e) Caminho mais curto pesado médio $\langle l \rangle$ em função do expoente de Hurst h com $d = 2$ (círculos) e $d = 3$ (triângulos) para o movimento browniano fracionário. Nos painéis (c)-(e), todas as curvas representam valores médios estimados a partir de um *ensemble* de 1000 séries para cada expoente de Hurst. As áreas sombreadas representam uma banda de um desvio padrão.

de todas essas séries em um procedimento de validação cruzada em cinco partes [54] para selecionar o melhor número k de primeiros vizinhos, o único hiperparâmetro do algoritmo. O restante das séries (25% do conjunto inicial), às quais o algoritmo nunca foi exposto, foram usadas para testar a precisão das predições. A Figura 2.7a mostra os valores verdadeiros e os valores preditos para o expoente de Hurst, na qual atingimos uma precisão de 97,7% medida pelo coeficiente de determinação⁷ R^2 . Esse resultado representa uma precisão muito boa considerando que as séries são relativamente curtas. Para efeitos de comparação, contrapomos o resultado da nossa abordagem baseada em redes ordinais com o resultado obtido da aplicação do método de análise de flutuação destendenciada⁸ (DFA) [55], um método confiável, robusto e tradicionalmente muito aplicado na estimativa do expoente de Hurst [56]. A Figura 2.7b mostra a relação entre os valores verdadeiros e os valores preditos de h resultantes da aplicação do método DFA a um subconjunto de 25% de amostras do mesmo *ensemble* de séries usado na avaliação do método das redes ordinais. Notamos imediatamente uma dispersão maior entre os valores verdadeiros e preditos, que pode ser quantificada por uma diminuição no coeficiente de determinação ($R^2 = 89,9\%$).

Comparamos também a precisão da abordagem de redes ordinais com outra transformação de série temporal em rede complexa, os grafos de quantil (Seção 1.2 do Capítulo 1) [9,35]. Construímos grafos de quantil a partir do mesmo conjunto de séries usados para as redes ordinais e testamos várias métricas de rede (grau médio, caminho mais curto pesado médio, diâmetro, coeficiente de aglomeração) como variáveis preditoras do expoente de Hurst. Verificamos que o grau médio foi a medida que produziu o melhor resultado nessa tarefa de regressão. A Figura 2.7c mostra a relação entre os valores verdadeiros e preditos para o expoente de Hurst usando grafos de quantil com 50 quantis. O coeficiente de determinação obtido foi $R^2 = 93,5\%$. Essa acurácia é maior que a do método DFA e menor que a precisão obtida usando redes ordinais, como mostra a Figura 2.6d.

O número de quantis q em um grafo de quantil tem um papel similar ao da *embedding dimension* d em uma rede ordinal visto que ambos os parâmetros definem o número de vértices nas redes. Assim, precisamos ajustar esses parâmetros para que tenhamos uma melhor comparação entre ambas as abordagens. Para esse fim, treinamos algoritmos de aprendizagem usando valores diferentes de d e q e estimamos os valores de R^2 no conjunto de teste para cada combinação. Observamos que o aumento de d diminui sistematicamente a qualidade da regressão para as redes ordinais e que o valor ótimo é $d = 2$ para essas séries temporais de 1024 observações. Por outro lado, grafos de quantil apresentam baixa acurácia quando o número de quantis é muito baixo ou muito alto. Sendo assim, os resultados da Figura 2.7 indicam que o desempenho da tarefa de regressão para valores otimizados de d

⁷Esse coeficiente é definido no contexto da avaliação de desempenho de algoritmos de aprendizagem estatística no Apêndice C.

⁸Do inglês, *detrended fluctuation analysis*. O procedimento de implementação desse método é detalhado no Apêndice D.

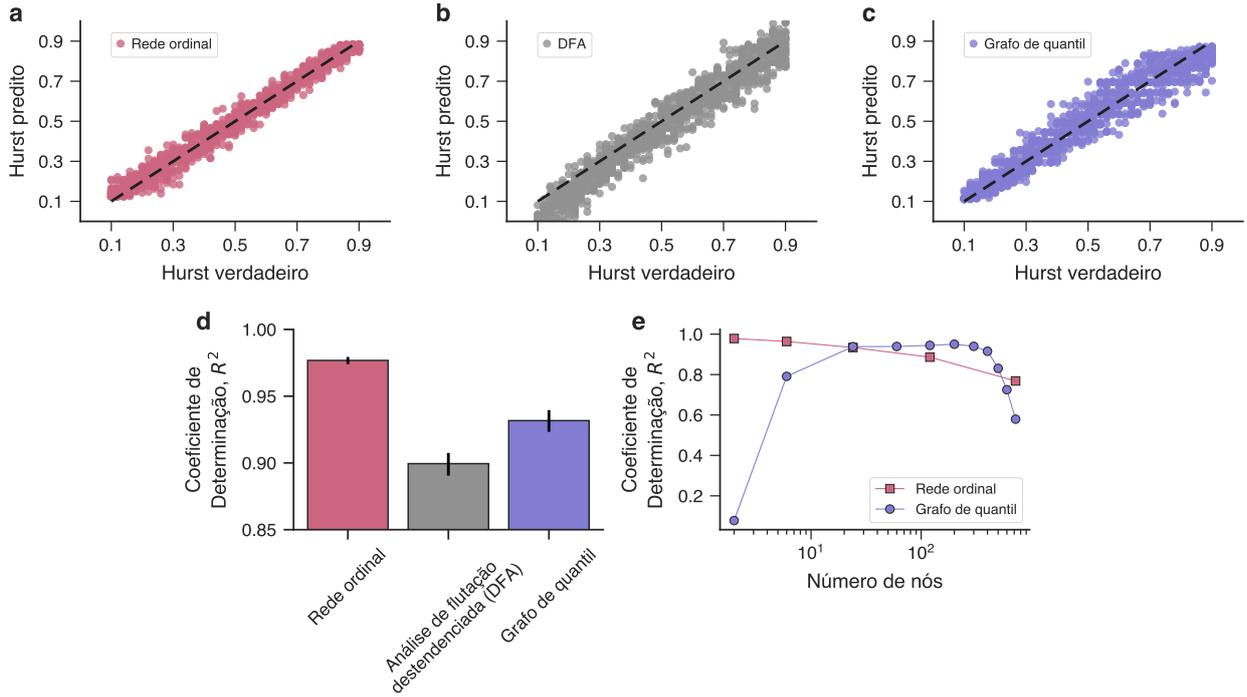


Figura 2.7: Estimativas do expoente de Hurst usando redes ordinais. Os painéis (a)-(c) mostram a relação entre os valores verdadeiros e preditos dos expoentes de Hurst obtidos via redes ordinais ($d = 2$), análise de flutuação destendenciada (DFA), e grafos de quantil ($q = 50$), respectivamente. As linhas tracejadas representam o resultado ideal. As previsões são obtidas aplicando o algoritmo regressor de k -primeiros vizinhos aos valores médios de caminho mais curto pesado $\langle l \rangle$ (redes ordinais) e grau (grafos de quantil) relacionados a séries de 1024 pontos de movimento browniano fracionário. Para o método DFA, o valor de h é obtido por um ajuste de mínimos quadrados entre a função de flutuação $F(s)$ e o parâmetro de escala s , ambos em escala logarítmica (isto é, $\log F(s) \propto h \log s$). (d) O gráfico de barras mostra a precisão de cada abordagem medida pelo coeficiente de determinação R^2 . Barras de erro representam intervalos de confiança de 95% obtidos pelo método de *bootstrapping* [54] em 1000 reamostras. (e) Dependência de R^2 com o número de vértices para redes ordinais (quadrados) e também para grafos de quantil (círculos).

e q é melhor para redes ordinais ($R^2 = 97,7\%$ para $d = 2$) do que para grafos de quantil ($R^2 = 93,5\%$ com $q = 50$).

Vale observar que o caminho mais curto pesado médio das redes ordinais pode ser interpretado como uma medida de persistência ou antipersistência. Esse fato é especialmente claro quando $d = 2$, pois os pesos das autoarestas ‘01’ \rightarrow ‘01’ e ‘10’ \rightarrow ‘10’ denotam a frequência de movimentos ascendentes e descendentes que ocorrem dentro de uma série. Logo, autoarestas desproporcionalmente pesadas indicam persistência. Caso as arestas conectando vértices distintos (‘01’ \rightarrow ‘10’ ou ‘10’ \rightarrow ‘01’) contenham a maior parte do peso distribuído na rede, temos uma indicação de antipersistência. Desse modo, valores maiores de $\langle l \rangle$ estão associados a uma prevalência das transições ‘01’ \rightarrow ‘10’ e ‘10’ \rightarrow ‘01’ enquanto valores meno-

res de $\langle l \rangle$ indicam uma frequência maior das transições ‘01’ \rightarrow ‘01’ e ‘10’ \rightarrow ‘10’ associadas às autoarestas.

2.4 Redes ordinais aplicadas a séries de atividade sísmica

Como última aplicação, passamos a investigar redes ordinais obtidas de séries temporais empíricas relativas à atividade sísmica terrestre. Em particular, analisamos séries temporais de magnitudes de terremotos da *Southern California Seismic Network* [57] entre 1990 e 2019. Mais especificamente, estudamos se as redes ordinais são capazes de detectar mudanças no comportamento das séries temporais devidas à ocorrência de um grande terremoto. Para tanto, selecionamos todos os eventos de magnitude maior que 7,0 e construímos duas séries temporais relativas a 200 eventos antes e depois de um grande terremoto. A Figura 2.8 mostra um exemplo de série de atividade sísmica (antes e depois) para o terremoto Baja California, de magnitude 7,2, ocorrido em 4 de abril de 2010 em Guadalupe Victoria, uma pequena cidade do estado de Baja California no México [58]. Temos ainda dois outros grandes terremotos: Landers (28 de junho de 1992) [59] de magnitude 7,3 e Hector Mine (16 de outubro de 1999) [60] de magnitude 7,1.

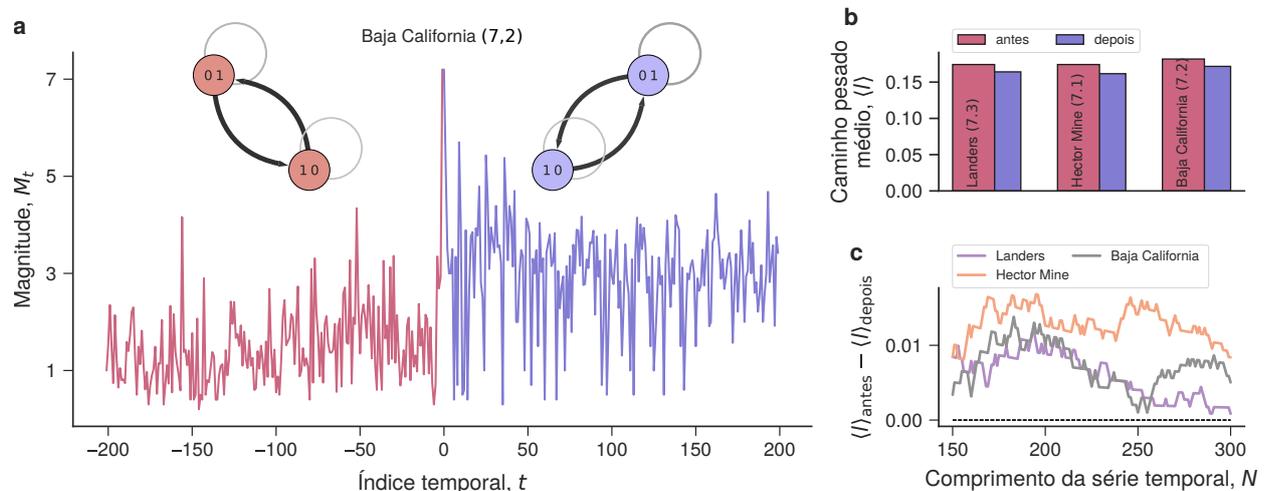


Figura 2.8: Detecção de mudanças na atividade sísmica terrestre usando redes ordinais. (a) Série temporal de magnitude de tremores antes (vermelho) e depois (azul) do grande terremoto Baja Califórnia [58]. As redes ordinais mostradas nesse painel foram obtidas usando as séries temporais de antes e depois do grande terremoto (com $d = 2$). (b) Caminho mais curto pesado médio $\langle l \rangle$ estimado a partir das redes ordinais construídas com séries temporais anteriores e posteriores a três grandes terremotos (indicados nas barras). Notamos que os valores de $\langle l \rangle$ diminuem após a ocorrência do *mainshock* nos três casos. (c) Cada curva colorida mostra a diferença entre o caminho mais curto médio antes ($\langle l \rangle_{\text{antes}}$) e depois ($\langle l \rangle_{\text{depois}}$) em função do número de eventos anteriores e posteriores ao *mainshock*. Observamos que $\langle l \rangle_{\text{antes}} - \langle l \rangle_{\text{depois}}$ é sempre positivo para N entre 150 e 300, o que demonstra a robustez desse resultado.

Mapeamos as séries temporais da atividade sísmica desses três grandes terremotos em redes ordinais com⁹ $d = 2$ (conforme ilustrado na Figura 2.8a) e calculamos a média do caminho mais curto pesado $\langle l \rangle$ para as redes antes ($\langle l \rangle_{\text{antes}}$) e depois ($\langle l \rangle_{\text{depois}}$) do terremoto. Os resultados obtidos (Figura 2.8b) mostram que os valores de $\langle l \rangle$ decrescem após a ocorrência de um grande terremoto. Verificamos que esse resultado se mantém quando variamos entre 150 e 300 o número de eventos sísmicos antes e depois do grande terremoto, conforme mostra a Figura 2.8c. Esse comportamento provavelmente está relacionado à lei de Omori [61], uma das leis sísmicas fundamentais¹⁰ que estabelece que o número de *aftershocks* por unidade de tempo decai como uma lei de potência do tempo decorrido após o *mainshock*. O decaimento da lei de Omori de alguma maneira pode introduzir correlações persistentes na série temporal após os maiores eventos, a qual pode ser a causa da redução do valor de $\langle l \rangle$.

⁹A escolha de *embedding dimension* $d = 2$ se justifica em razão do tamanho dessas séries temporais.

¹⁰As leis fundamentais da dinâmica sísmica terrestre são [62]: (1) Lei de Gutenberg-Richter, que descreve a distribuição de probabilidade de ocorrência de terremotos em função da magnitude ou energia; (2) Lei de Omori, que descreve a dinâmica de decaimento da produção de tremores secundários (*aftershocks*) após um tremor principal (*mainshock*); (3) Lei que rege a estatística de intervalos de tempo entre tremores consecutivos de magnitude superior a um limiar predefinido; (4) Lei de produtividade, que relaciona a energia de um *mainshock* ao número de *aftershocks* desencadeados; (5) Lei de Bâth, que afirma que a diferença média entre a magnitude de um *mainshock* e o maior *aftershock* é igual a 1,2.

Conclusões e Perspectivas

Neste trabalho, apresentamos algumas das principais propostas para transformações de séries temporais em redes complexas encontradas na literatura: grafos de visibilidade, grafos de quantil, redes de recorrência e redes ordinais. Para tanto, descrevemos cada uma dessas propostas e reproduzimos aplicações relativas a séries de natureza caótica e estocástica. Esses algoritmos fazem parte do recém-surgido campo de pesquisa de redes de séries temporais cujos métodos têm despertado considerável interesse e encontrado aplicações em séries temporais de diversas naturezas.

Também apresentamos uma investigação acerca de redes ordinais mapeadas de séries temporais de natureza estocástica, uma vez que um estudo desse tipo faltava à literatura científica. Em particular, analisamos redes ordinais mapeadas de ruído branco, sinais periódicos ruidosos, movimento browniano fracionário e séries temporais de atividade sísmica terrestre. Oferecemos uma descrição detalhada das redes ordinais mapeadas de ruído branco (redes ordinais aleatórias), revelando algumas propriedades contraintuitivas como a distribuição não uniforme dos pesos das arestas e a existência de autoarestas somente em nós associados a permutações ascendentes ou descendentes. Também propomos um procedimento para a construção da forma exata da matriz de adjacência dessas redes ordinais aleatórias a fim de usá-las na detecção de comportamento não aleatório em séries temporais. Para sinais periódicos ruidosos, nossos resultados indicaram que a entropia global dos vértices estimada a partir das redes ordinais é mais robusta na presença de ruído que a entropia de permutação. Ainda, demonstramos a utilidade das redes ordinais na estimativa do expoente de Hurst de séries temporais e na detecção de mudanças em séries de atividade sísmica após a ocorrência de grandes tremores.

Além dos resultados relativos às redes ordinais apresentados nesta dissertação, temos como perspectiva a realização de uma investigação mais detalhada acerca do processo de mapeamento inverso de redes ordinais, isto é, da obtenção de uma réplica da série temporal

a partir de sua rede ordinal. Tal como no caso dos grafos de quantil, esse procedimento passa pela realização de uma caminhada aleatória sobre a rede. Imaginamos, por exemplo, que capturando os padrões ordinais ao longo de um passeio aleatório sobre uma rede ordinal mapeada de uma amostra de ruído gaussiano fracionário possamos reproduzir uma réplica dessa série de mesmo expoente de Hurst que a série original.

Outras possibilidades de pesquisas originais incluem a análise de redes ordinais mapeadas de séries fractais (motivada pelo estudo sobre o movimento browniano fracionário) e a extensão natural do método de redes ordinais para a transformação de dados de maior dimensão (como imagens) em redes para fins de caracterização.

A.1 Redes complexas e sua representação

Uma rede (ou grafo) é um conjunto de vértices conectados por arestas (Figura A.1) [1, 2]. Os vértices, também chamados nós ou atores, usualmente representam elementos que constituem um determinado sistema, enquanto as arestas, também chamadas de ligações ou laços, denotam algum tipo de relação ou interação entre esses constituintes [1, 2]. Devido à simplicidade e generalidade dessa ideia de representação, são inúmeros os possíveis exemplos de sistemas que podem ser aproximados por redes e estudados sob essa perspectiva. Alguns exemplos são redes sociais, a internet, redes neuronais, redes de colaboração e citação em comunidades científicas entre outros [1, 2, 4].

Em sua forma mais comum, as redes têm diferentes pares de vértices ligados por apenas uma aresta (Figura A.1a). Nesse caso, são chamadas de grafos simples, em oposição a outros tipos de grafos que podem apresentar relações mais complexas entre seus vértices como autoarestas (um nó conectando-se a si mesmo) ou múltiplas arestas (Figura A.1b) [1, 2]. Para além da mera relação binária da ocorrência da aresta, outras propriedades, como o peso e a direção da ligação também são importantes. Redes nas quais as arestas possuem peso e direção são chamadas de redes direcionadas e pesadas (Figura A.1c).

Matematicamente, um grafo é completamente definido por uma matriz chamada matriz de adjacência [1, 2]. Para um grafo simples, sua matriz de adjacência \mathbf{A} tem o elemento a_{ij} igual a 1 caso os vértices i e j estejam conectados na rede e 0 caso o contrário. A matriz de

adjacência do grafo mostrado na Figura A.1a, por exemplo, é

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}. \quad (\text{A.1})$$

Para redes pesadas, os elementos da matriz de adjacência são, em geral, diferentes de 0 e 1. Em redes direcionadas, caso o elemento a_{ij} da matriz de adjacência \mathbf{A} seja não nulo, existe uma ligação direcionada do i -ésimo para o j -ésimo vértice da rede. A matriz de adjacência da rede direcionada e pesada mostrada na Figura A.1c, por exemplo, é

$$\mathbf{A} = \begin{pmatrix} 0 & 2 & 0 & 2 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}. \quad (\text{A.2})$$

Uma vez que temos uma representação matemática conveniente de um grafo, podemos calcular algumas métricas da rede a fim de caracterizar sua estrutura de conexões [1, 2, 4]. Isso é importante porque as redes encontradas no mundo real são frequentemente chamadas de redes complexas por apresentar uma disposição de arestas entre a completa regularidade e a completa aleatoriedade [2]. Métricas como caminho mais curto l , diâmetro L , coeficiente de aglomeração \mathcal{C} e grau k são algumas das medidas mais comuns [1, 2, 4].

O caminho mais curto (caminho geodésico) $l_{i,j}$ entre dois vértices i e j é o menor número de arestas que se atravessadas levam de um dos vértices ao outro [1, 2]. O valor médio do caminho mais curto sobre todos os pares de vértices da rede \mathcal{L} é uma medida importante associada ao fenômeno de “mundo pequeno” observado em redes complexas [32]. Em redes direcionadas, a determinação do caminho mais curto entre dois vértices tem de respeitar as direções das arestas. Em redes pesadas, o caminho mais curto passa a ser o caminho de menor soma dos pesos das arestas. Por sua vez, o diâmetro L de uma rede é o maior caminho mais curto entre dois vértices quaisquer da rede [1, 2].

O coeficiente de aglomeração¹ \mathcal{C} é uma medida que avalia a formação de triângulos em redes [1, 2, 32]. Definindo o coeficiente de aglomeração local C_i de um vértice i como a fração do número de conexões observadas entre os primeiros vizinhos de i e o número de conexões possíveis entre esses vizinhos, o coeficiente de aglomeração é calculado tomando a média dessa medida local para todos os vértices da rede. Redes complexas costumam apresentar

¹O coeficiente de aglomeração usado aqui é devido a Watts e Strogatz [32]. Para outras definições veja a referência [1].

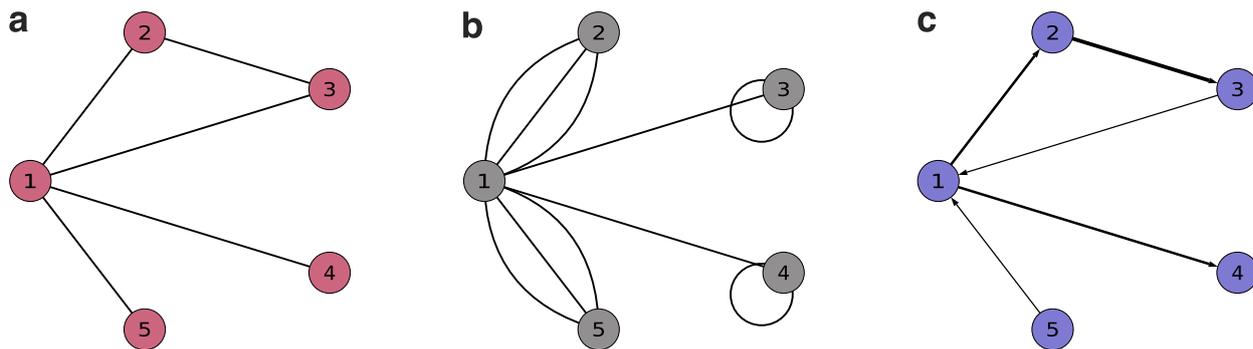


Figura A.1: Ilustração de diferentes tipos de grafos. (a) Grafo simples. (b) Grafo (não simples) apresentando multiarestas e autoarestas. (c) Rede direcionada e pesada.

valores grandes para esse coeficiente quando comparadas a redes formadas pela inserção aleatória de arestas entre vértices [1, 32].

Por fim, o grau k de um vértice é simplesmente o número de arestas conectadas a ele [1, 2]. Em redes direcionadas, um vértice tem duas medidas de grau: o grau de entrada k_{in} (número de arestas que chegam ao vértice) e o grau de saída k_{out} (número de arestas que saem do vértice). Entretanto, além do grau de um vértice particular da rede, é interessante conhecer a distribuição de grau da rede, isto é, a probabilidade $p(k)$ de encontrarmos um vértice conectado a k outros vértices da rede. Em redes complexas, é comum observarmos distribuições de grau de cauda pesada associadas à presença de vértices muito conectados na rede, os chamados *hubs* [25].

A.2 Modelos de redes complexas

Uma vez que representamos convenientemente um grafo e quantificamos sua estrutura usando algumas das métricas desenvolvidas no estudo de redes, o próximo passo é a construção de modelos teóricos que reproduzam características observadas em redes do mundo real. Sendo assim, passamos a descrever três dos principais modelos usados no estudo de redes complexas: Erdős-Rényi [63], Watts-Strogatz [32] e Barabási-Albert [25].

O modelo de grafos aleatórios ou modelo de Erdős-Rényi (Figura A.2a) [63] é um modelo estocástico de formação de grafos. Em uma das descrições desse modelo, o número de vértices n da rede é fixado e uma aresta entre um par qualquer de vértices existe com probabilidade p [1, 2, 4]. É importante notar que esse modelo não gera apenas uma rede, mas um *ensemble* de grafos simples no qual cada membro desse conjunto é uma realização particular do modelo. Propriedades típicas de um grafo aleatório são então avaliadas a partir de médias sobre todo o *ensemble*.

Denotando por G um grafo desse *ensemble*, a probabilidade de obtenção dessa realização

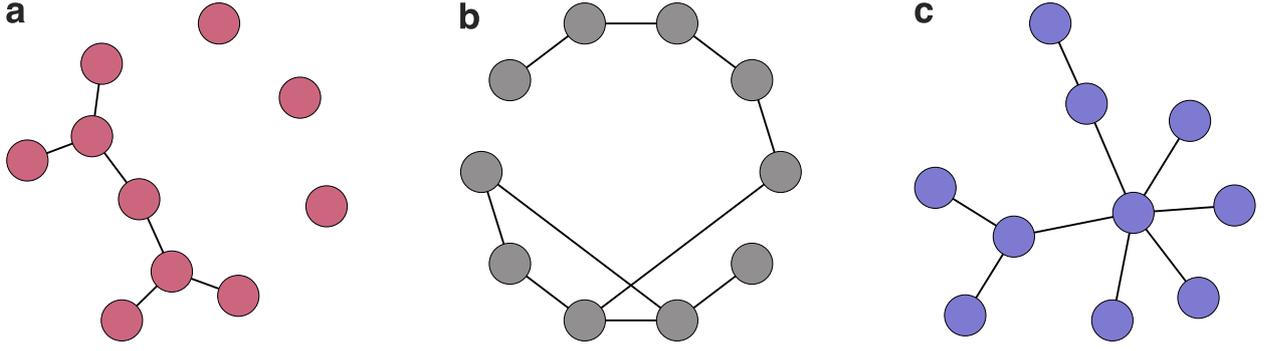


Figura A.2: Ilustração de algumas redes obtidas de modelos de redes complexas.

(a) Realização do modelo de Erdős-Rényi com os parâmetros $n = 10$ e $p = 0,1$. (b) Realização do modelo de Watts-Strogatz com probabilidade $p = 0,1$ de rearranjo de cada aresta de uma rede regular com $k = 2$. (c) Realização do modelo de Barabási-Albert com cada passo adicionando um vértice ligado a uma aresta à rede.

particular do modelo é dada pela distribuição de Bernoulli [1, 2]

$$p(G) = p^m(1-p)^{\binom{n}{2}-m}, \quad (\text{A.3})$$

na qual m denota o número de arestas e $\binom{n}{2}$ é o número de diferentes pares de vértices na rede. Além disso, a probabilidade $p(m)$ de se obter um grafo com m arestas é dada pelo produto entre $p(G)$ e o número de maneiras de dispor m arestas entre todos os pares de vértices [1], ou seja,

$$p(m) = \binom{\binom{n}{2}}{m} p^m (1-p)^{\binom{n}{2}-m}. \quad (\text{A.4})$$

Usando esse resultado, podemos calcular o número típico de arestas para um grafo aleatório:

$$\begin{aligned} \langle m \rangle &= \sum_{m=0}^{\binom{n}{2}} m p(m) \\ &= \sum_{m=0}^{\binom{n}{2}} m \binom{\binom{n}{2}}{m} p^m (1-p)^{\binom{n}{2}-m} \\ &= \binom{n}{2} p \sum_{m=1}^{\binom{n}{2}} \frac{(\binom{n}{2}-1)!}{((\binom{n}{2}-1)-(m-1))!(m-1)!} p^{(\binom{n}{2}-1)} p^{(\binom{n}{2}-1)-(m-1)} \quad (\text{A.5}) \\ &= \binom{n}{2} p \sum_{m=1}^{\binom{n}{2}} \binom{\binom{n}{2}-1}{m-1} p^{(\binom{n}{2}-1)} p^{(\binom{n}{2}-1)-(m-1)} \\ &= \binom{n}{2} p. \end{aligned}$$

A distribuição de grau $p(k)$ de um grafo aleatório pode ser obtida considerando que a

probabilidade de que um determinado vértice esteja ligado a k dos $n - 1$ outros vértices do grafo é uma distribuição de Bernoulli. Tendo em mente que existem $\binom{n-1}{k}$ maneiras diferentes de escolher k entre $n - 1$ vértices [1], a distribuição de grau de um grafo aleatório é também uma distribuição Binomial dada por

$$p(k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}. \quad (\text{A.6})$$

Consequentemente, o valor médio de k sobre todo o *ensemble* é

$$\langle k \rangle = \sum_{k=0}^{n-1} k p(k) = (n-1)p. \quad (\text{A.7})$$

A partir dessa última relação, fixando $\langle k \rangle$ de modo que p varie em função do número de vértices n e tomando o limite de $n \rightarrow \infty$, podemos usar as aproximações [1]

$$\binom{n-1}{k} = \frac{(n-1)!}{(n-1-k)!k!} \cong \frac{(n-1)^k}{k!} \quad (\text{A.8})$$

e

$$\begin{aligned} \ln(1-p)^{n-1-k} &= (n-1-k) \ln(1-p) \\ &= -(n-1-k)p \\ &= -\frac{(n-1-k)}{(n-1)} \langle k \rangle \\ &\sim -\langle k \rangle, \end{aligned} \quad (\text{A.9})$$

uma vez que a Eq. A.7 implica em $p \rightarrow 0$ nesse limite. Assim, a distribuição de grau do modelo de Erdős-Rényi é aproximada pela distribuição de Poisson

$$p(k) = \frac{e^{-\langle k \rangle} \langle k \rangle^k}{k!}. \quad (\text{A.10})$$

Por esse motivo, esses grafos também são chamados de grafos de Poisson [1, 2].

Outra medida que pode ser calculada é o coeficiente de aglomeração \mathcal{C} . Interpretando o coeficiente de aglomeração local C_i de um vértice i da rede como a probabilidade de conexão entre dois primeiros vizinhos desse vértice e notando que essa probabilidade é igual a p (as arestas ocorrem independentemente umas das outras), temos $C_i = p$ para qualquer vértice i na rede e, logo, o coeficiente de aglomeração é

$$\mathcal{C} = p = \frac{\langle k \rangle}{n-1}. \quad (\text{A.11})$$

É interessante notar que se $\langle k \rangle$ é mantido constante, tomando o limite de $n \rightarrow \infty$ esse

coeficiente vai a zero.

Uma última característica importante do modelo de grafos aleatórios é a dependência do valor médio do caminho mais curto com o número de vértices do grafo. Essa dependência é da forma [1, 2, 4]

$$\mathcal{L} \sim \frac{\ln n}{\ln \langle k \rangle}, \quad (\text{A.12})$$

em que o crescimento logarítmico marca a propriedade de mundo pequeno [2].

Como mencionamos, redes complexas frequentemente apresentam a propriedade de mundo pequeno, valores muito maiores para o coeficiente de aglomeração que os preditos pela Eq. A.11 e distribuição de grau de cauda pesada [1, 2, 4]. Como podemos ver, o modelo de Erdős-Rényi falha em reproduzir simultaneamente essas características.

Outro modelo de redes complexas, o modelo de “mundo pequeno” ou modelo de Watts-Strogatz [32], representa uma tentativa de preservar a propriedade de mundo pequeno, observada nos grafos aleatórios, e simultaneamente produzir valores grandes para o coeficiente de aglomeração [32]. Uma rede de mundo pequeno (Figura A.2b) é obtida a partir de uma rede regular na forma de um anel² com n vértices ligados aos seus k primeiros vizinhos (sendo k um número par) na qual cada uma dessas arestas é rearranjada com probabilidade p . No limite em que $p \rightarrow 0$, o grafo mantém sua estrutura de rede cristalina (alto coeficiente de aglomeração) enquanto $p \rightarrow 1$ produz um grafo aleatório (propriedade de mundo pequeno). As características de mundo pequeno e alta formação de triângulos são simultaneamente encontradas para pequenos valores de p ($p < 0,1$). Para esses valores, o modelo apresenta uma acentuada diminuição no caminho mais curto médio \mathcal{L} , enquanto o coeficiente de aglomeração \mathcal{C} da rede permanece praticamente inalterado [32].

Por fim, o modelo de Barabási-Albert [25] tem como principal particularidade sua distribuição de grau de cauda pesada. Nesse modelo, partindo de uma rede inicial, a cada passo é adicionado um novo vértice à rede e um número fixo de arestas entre esse vértice e os vértices já presentes na rede. Além disso, as arestas são traçadas com probabilidade proporcional ao grau do vértice na rede, de modo que os vértices mais conectados tendem a aumentar ainda mais suas ligações em uma espécie de efeito no qual o “rico fica mais rico” [25]. Barabási e Albert [25] encontraram que as redes geradas segundo esse mecanismo apresentam uma distribuição de grau da forma $p(k) \sim k^{-\gamma}$ com $\gamma = 3$. Devido à forma dessa distribuição, essas redes são também chamadas de redes livres de escala [2, 4].

²Rede cristalina unidimensional com condição periódica de contorno.

B.1 Séries temporais, autocorrelações e processos estocásticos

Uma série temporal $\{x_t\}_{t=1,\dots,N}$ é um conjunto de observações de uma variável feitas sequencialmente no tempo [64]. Exemplos comuns de séries temporais acontecem em economia (séries históricas de indicadores econômicos) e em ciências físicas como meteorologia (séries climáticas) e geofísica (séries de atividade sísmica).

Em virtude da complexidade comumente associada às séries temporais desses sistemas naturais e sociais, seu estudo é geralmente realizado a partir de uma perspectiva probabilística. Entretanto, enquanto experimentos canônicos em probabilidade (como lançamentos de moedas ou dados) assumem de partida que os resultados de duas realizações consecutivas do mesmo experimento são independentes entre si, uma característica que marca o estudo de séries temporais é a presença de dependência entre resultados de observações sucessivas [64].

Em estatística, medidas de correlação são frequentemente usadas para avaliar associação ou dependência entre duas variáveis. Supondo, por exemplo, dois conjuntos de observações $\{x_1, x_2, \dots, x_N\}$ e $\{y_1, y_2, \dots, y_N\}$ de grandezas distintas x e y , podemos definir o coeficiente de correlação de Pearson R [64] como

$$R = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}}, \quad (\text{B.1})$$

sendo $\bar{x} = \sum_{i=1}^N x_i/N$ e $\bar{y} = \sum_{i=1}^N y_i/N$ as respectivas médias desses conjuntos de observa-

ções. Do mesmo modo, podemos generalizar essa medida de correlação com o objetivo de avaliar associações entre pares de observações consecutivas em uma série temporal. Para tanto, dada uma série temporal $\{x_t\}_{t=1,\dots,N}$, passamos a considerar os pares de observações $(x_1, x_2), (x_2, x_3), \dots, (x_{N-1}, x_N)$. Em seguida, definimos o coeficiente de correlação serial ou, simplesmente, autocorrelação [64] como

$$r = \frac{\sum_{i=1}^{N-1} (x_i - \bar{x}^{(1)})(x_{i+1} - \bar{x}^{(2)})}{\sqrt{\sum_{i=1}^{N-1} (x_i - \bar{x}^{(1)})^2 \sum_{i=2}^N (x_{i+1} - \bar{x}^{(2)})^2}}, \quad (\text{B.2})$$

com $\bar{x}^{(1)} = \sum_{i=1}^{N-1} x_i / (N-1)$ e $\bar{x}^{(2)} = \sum_{i=2}^N x_i / (N-1)$. Caso a série temporal seja constituída de um número razoavelmente grande de observações podemos supor $\bar{x}^{(1)} = \bar{x}^{(2)} \sim \bar{x}$ e aproximar ambos os termos no radical da expressão anterior para obter

$$r = \frac{\sum_{i=1}^{N-1} (x_i - \bar{x})(x_{i+1} - \bar{x})}{\sum_{i=1}^N (x_i - \bar{x})^2}. \quad (\text{B.3})$$

Considerando que a dependência entre elementos de uma série temporal pode se estender a pares de observações não consecutivas, podemos generalizar a medida anterior de autocorrelação definindo a autocorrelação de *lag* k [64] como

$$r(k) = \frac{\sum_{i=1}^{N-k} (x_i - \bar{x})(x_{i+k} - \bar{x})}{\sum_{i=1}^N (x_i - \bar{x})^2}. \quad (\text{B.4})$$

Essa definição de autocorrelação representa uma maneira geral de estimarmos correlações em uma série temporal. Tendo isso em mente, a tarefa de estimativa é simplificada expressando a autocorrelação em termos de outra quantidade chamada autocovariância. A autocovariância de *lag* k [64] pode ser definida por

$$c(k) = \frac{1}{N} \sum_{i=1}^{N-k} (x_i - \bar{x})(x_{i+k} - \bar{x}) \quad (\text{B.5})$$

e, conseqüentemente, a autocorrelação de *lag* k pode ser reescrita como

$$r(k) = \frac{c(k)}{c(0)}. \quad (\text{B.6})$$

Estimativas empíricas de interdependência feitas a partir dessa equação são importantes na seleção de modelos probabilísticos apropriados à descrição de séries temporais. Coletivamente, esses modelos probabilísticos são chamados de processos estocásticos [64].

Um processo estocástico (aleatório), denotado por $\{X(t), t \in T\}$, é uma coleção de variáveis aleatórias $X(t)$ indexadas pela variável t que, usualmente, representa o tempo [65]. Caso o conjunto T dos índices seja um conjunto enumerável, o processo é dito ser discreto [65].

Caso T seja um intervalo da reta real, o processo é contínuo [65]. Nessa abordagem probabilística, uma série temporal $\{x(t), t \in T\}$ pode ser pensada como uma realização particular de um processo estocástico [64].

Um processo estocástico pode ser definido pela distribuição de probabilidade conjunta das variáveis aleatórias $X(t_1), X(t_2), \dots, X(t_N)$ [66]. Para qualquer $n \geq 1$, podemos escrever a distribuição acumulada

$$F(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) = P(X(t_1) \leq x_1, X(t_2) \leq x_2, \dots, X(t_n) \leq x_n), \quad (\text{B.7})$$

na qual os índices t_1, \dots, t_n representam instantes de amostragem do processo estocástico. Por outro lado, um processo aleatório também pode ser especificado pelos momentos dessa distribuição [66]. Mais especificamente, os momentos da distribuição são

$$m_q(q_1, q_2, \dots, q_n; t_1, t_2, \dots, t_n) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} x_1^{q_1} x_2^{q_2} \dots x_n^{q_n} f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n, \quad (\text{B.8})$$

com $q = \sum_{i=1}^n q_i$ representando a ordem do momento e $f(x_1, x_2, \dots, x_n)$ representando a função densidade de probabilidade do processo aleatório [66]. Entretanto, notada a complexidade de determinação tanto da distribuição (Eq. B.7) quanto dos momentos (Eq. B.8), é comum avaliarmos apenas os primeiros momentos da distribuição, notadamente, média, variância e autocovariância [64, 66].

A média e variância de um processo estocástico são definidas como

$$\mu(t) = \mathbf{E}\{X(t)\} \quad (\text{B.9})$$

e

$$\gamma(t) = \mathbf{Var}\{X(t)\} = \mathbf{E}\{[X(t) - \mu(t)]^2\}, \quad (\text{B.10})$$

enquanto a função de autocovariância é expressa por

$$\begin{aligned} \gamma(t_i, t_j) &= \mathbf{Cov}\{X(t_i), X(t_j)\} \\ &= \mathbf{E}\{[X(t_i) - \mu(t_i)][X(t_j) - \mu(t_j)]\}. \end{aligned} \quad (\text{B.11})$$

Uma importante distinção entre processos estocásticos diz respeito à estacionariedade do processo. Um processo estocástico é dito estritamente estacionário se as distribuições de probabilidade de $X(t_1), X(t_2), \dots, X(t_n)$ e $X(t_1 + \tau), X(t_2 + \tau), \dots, X(t_n + \tau)$ são as mesmas [66], ou seja,

$$F(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) = F(x_1, x_2, \dots, x_n; t_1 + \tau, t_2 + \tau, \dots, t_n + \tau) \quad (\text{B.12})$$

para quaisquer t_1, \dots, t_n e τ em T [66]. Especificamente, para o primeiro e segundo momento

do processo estocástico, estacionariedade estrita implica em valores constantes de média e variância para o processo [64], isto é,

$$\mu(t) = \mathbf{E}\{X(t)\} = \mu \quad (\text{B.13})$$

e

$$\gamma(t) = \mathbf{E}\{X^2(t)\} - \mu^2 = \sigma^2, \quad (\text{B.14})$$

sendo μ e σ constantes. A função de autocovariância, por sua vez, passa a ser função apenas do *lag* k entre duas observações [64] e tem a forma

$$\begin{aligned} \gamma(k) &= \mathbf{Cov}\{X(t), X(t+k)\} \\ &= \mathbf{E}\{X(t)X(t+k)\} - \mu^2. \end{aligned} \quad (\text{B.15})$$

Outra definição possível de estacionariedade, menos restritiva, é chamada de estacionariedade fraca ou de segunda ordem. Um processo estocástico é fracamente estacionário ou estacionário de segunda ordem caso sua média seja constante, sua variância finita e sua função de autocovariância dependa somente do *lag* entre duas observações [64,66]. Matematicamente, essas condições são expressas por

$$\mu(t) = \mathbf{E}\{X(t)\} = \mu, \quad (\text{B.16})$$

$$\mathbf{E}\{X^2(t)\} < \infty \quad (\text{B.17})$$

e

$$\gamma(k) = \mathbf{E}\{X(t)X(t+k)\} - \mu^2. \quad (\text{B.18})$$

Um detalhe interessante acerca das definições anteriores é que estacionariedade estrita não implica necessariamente em estacionariedade fraca, pois um processo estocástico pode ser estritamente estacionário e não satisfazer $\mathbf{E}\{X^2(t)\} < \infty$ [66]. Também, estacionariedade estrita resulta de uma condição imposta à distribuição de probabilidade conjunta que caracteriza o processo aleatório (Eq. B.11), enquanto estacionariedade fraca ou de segunda ordem resulta de restrições nos momentos dessa distribuição [64,66]. Uma classe particular de processos estocásticos para os quais estacionariedade de segunda ordem implica em estacionariedade estrita, são os processos estocásticos gaussianos ou normais [64,66]. Esses processos recebem esse nome porque a distribuição de probabilidade conjunta de $X(t_1), X(t_2), \dots, X(t_N)$ é normal.

Estabelecida a estacionariedade de um processo aleatório, é interessante investigar a extensão de suas autocorrelações, isto é, investigar se as correlações entre as variáveis aleatórias que constituem o processo apresentam um tempo típico de decaimento. A função de autocorrelação teórica de um processo estocástico é definida, similarmente à sua equivalente

empírica (Eq. B.6), como

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)}. \quad (\text{B.19})$$

A investigação acerca da extensão da interdependência entre variáveis do processo passa por analisar a convergência da integral da função de autocorrelação [67], isto é,

$$\int_0^\infty \rho(k) dk. \quad (\text{B.20})$$

Caso essa integral convirja, existe uma escala típica de tempo em que as variáveis são correlacionadas e o processo estocástico é dito apresentar correlações de curto alcance ou memória de curto prazo [67]. Uma função de autocorrelação $\rho(k) = e^{-k/\tau_c}$ tem a constante τ_c como medida característica do tempo de decaimento das autocorrelações. Caso a integral divirja, não existe uma constante que representa uma escala típica de tempo para as correlações e o processo estocástico é dito apresentar correlações de longo alcance ou memória de longo prazo [67]. Funções de autocorrelação tipo lei de potência $\rho(k) \sim k^{-\alpha}$, sendo $0 < \alpha \leq 1$ o expoente da lei de potência, são típicas de processos com correlações de longo alcance.

B.2 Movimento browniano fracionário

O movimento browniano fracionário $B_h(t)$ é um processo estocástico contínuo, autossimilar¹ e não estacionário introduzido por Mandelbrot e Van Ness [24]. Esse processo aleatório pode ser representado pela integral estocástica

$$B_h(t) = B_h(0) + \frac{1}{\Gamma(h + \frac{1}{2})} \left\{ \int_{-\infty}^0 [(t-s)^{h-\frac{1}{2}} - (-s)^{h-\frac{1}{2}}] dB(s) + \int_0^t (t-s)^{h-\frac{1}{2}} dB(s) \right\}, \quad (\text{B.21})$$

na qual o integrando $B(s)$ é o movimento browniano usual, $\Gamma(x) = \int_0^\infty z^{x-1} e^{-z} dz$ é a função gama e $h \in (0, 1)$ é o parâmetro ou expoente de Hurst [24, 69].

Os incrementos (diferenças) do movimento browniano fracionário $W_h(t) = B_h(t+1) - B_h(t)$ definem um processo estocástico estacionário discreto chamado de ruído gaussiano fracionário [68, 69]. Esse ruído é um processo gaussiano, com cada uma das variáveis aleatórias $W_h(t_1), W_h(t_2), \dots$ distribuídas de acordo com uma normal padrão. A função de autocovariância desse processo estacionário é dada por [68, 69]

$$\gamma(k) = \frac{1}{2} (|k-1|^{2h} - 2|k|^{2h} + |k+1|^{2h}), \quad (\text{B.22})$$

¹Por autossimilaridade nos referimos a igualdade em distribuição entre $B_h(ct)$ e $c^h B_h(t)$ para $c > 0$. O parâmetro h (expoente de Hurst) é o parâmetro de autossimilaridade do processo [68, 69].

com k representando a defasagem (*lag*) entre duas observações de $W_h(t)$.

A depender do valor do expoente de Hurst, o ruído gaussiano fracionário pode ser correlacionado de curto alcance, não correlacionado ou correlacionado de longo alcance [68, 69]. Para evidenciar esses diferentes comportamentos, passamos a analisar o comportamento da função $\gamma(k)$ para valores grandes de k . Notamos, primeiramente, que usando $g(x) = (1 - x)^{2h} - 2 + (1 + x)^{2h}$ podemos reescrever a função de autocovariância como $\gamma(k) = \frac{1}{2}k^{2h}g(1/k)$ caso $k \geq 1$ [68]. Sendo assim, analisar o comportamento da função g para pequenos valores de x é o mesmo que entender $\gamma(k)$ quando k cresce. Usando aproximações binomiais de segunda ordem em g encontramos

$$\begin{aligned} g(x) &\sim 1 - 2hx + h(2h - 1)x^2 - 2 + 1 + 2hx + h(2h - 1)x^2 \\ &\sim 2h(2h - 1)x^2, \end{aligned} \tag{B.23}$$

quando $x \rightarrow 0$. Logo, podemos aproximar $\gamma(k) = \frac{1}{2}k^{2h}g(1/k)$ no limite $k \rightarrow \infty$ por

$$\gamma(k) \sim h(2h - 1)k^{2h-2}. \tag{B.24}$$

A expressão anterior é positiva para valores de $h > \frac{1}{2}$ e negativa para $h < \frac{1}{2}$, indicando que o ruído fracionário gaussiano pode ser positiva ou negativamente correlacionado, respectivamente. As correlações se anulam quando $h = \frac{1}{2}$, recuperando a independência entre os incrementos de um movimento browniano usual. Ainda mais, para $h > \frac{1}{2}$, o expoente $2h - 2$ da forma assintótica da função de autocovariância é menor que um e a integral da função de autocorrelação não converge, tornando o processo correlacionado de longo alcance. Para $h < \frac{1}{2}$, essa integral converge e o processo é anticorrelacionado de curto alcance.

B.3 Simulação do movimento browniano fracionário

Antes de descrevermos o procedimento que usamos na simulação de séries temporais (amostras) de movimento browniano fracionário, apresentamos a generalização da distribuição normal em n dimensões, também chamada de normal multivariada. Uma distribuição normal multivariada é uma distribuição da forma [70]

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \right\}, \tag{B.25}$$

na qual os vetores coluna n -dimensionais $\mathbf{x} = (x_1, x_2, \dots, x_n)$ e $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_n)$ representam vetores de variáveis e médias, respectivamente. A matriz $\boldsymbol{\Sigma}$, chamada matriz de covariância, é responsável pela interdependência entre as variáveis. Sua inversa $\boldsymbol{\Sigma}^{-1}$ é chamada de matriz de precisão. Assim como uma distribuição normal unidimensional está associada a uma única variável aleatória X , uma distribuição normal multivariada representa a

distribuição conjunta de um vetor ou uma coleção de variáveis aleatórias (X_1, X_2, \dots, X_n) . Comparando com o caso unidimensional em que uma distribuição normal é caracterizada por sua média e variância, a expressão anterior mostra que uma normal multivariada é definida pelo vetor de médias e pela matriz de covariância.

Apresentada a distribuição normal multivariada, passamos a demonstrar a forma de suas distribuições marginal e condicional, com base na referência [71]. Primeiramente, escrevemos o vetor de variáveis \mathbf{x} como

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix}, \quad (\text{B.26})$$

sendo \mathbf{x}_1 um vetor p -dimensional e \mathbf{x}_2 um vetor de dimensão q , com $p+q = n$. Analogamente, reescrevemos as estruturas de média e covariância como

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad (\text{B.27})$$

com os vetores $\boldsymbol{\mu}_1$ e $\boldsymbol{\mu}_2$ de mesmas dimensões que \mathbf{x}_1 e \mathbf{x}_2 , respectivamente, e

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}. \quad (\text{B.28})$$

Os blocos $\boldsymbol{\Sigma}_{11}$ e $\boldsymbol{\Sigma}_{12}$ da matriz de covariância são matrizes quadradas de ordens p e q , enquanto $\boldsymbol{\Sigma}_{21}$ e $\boldsymbol{\Sigma}_{22}$ têm dimensões $q \times p$ e $p \times q$. Devido à simetria da matriz de covariância, $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}^T$. A matriz de precisão, escrita em blocos, é representada por

$$\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{\Sigma}^{11} & \boldsymbol{\Sigma}^{12} \\ \boldsymbol{\Sigma}^{21} & \boldsymbol{\Sigma}^{22} \end{pmatrix}, \quad (\text{B.29})$$

com cada um de seus blocos $(\boldsymbol{\Sigma}^{11}, \boldsymbol{\Sigma}^{12}, \boldsymbol{\Sigma}^{21}, \boldsymbol{\Sigma}^{22})$ tendo as mesmas dimensões dos respectivos blocos de $\boldsymbol{\Sigma}$.

Essas diferentes representações de \mathbf{x} , $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ nos ajudam a demonstrar os seguintes resultados:

1. As distribuições marginais, $f(\mathbf{x}_1)$ ou $f(\mathbf{x}_2)$, de uma normal multivariada $f(\mathbf{x})$ são, também, normais multivariadas;
2. A distribuição condicional $f(\mathbf{x}_1|\mathbf{x}_2)$ de \mathbf{x}_1 dado \mathbf{x}_2 é uma normal multivariada.

Para demonstrar esses resultados, reescrevemos o expoente $(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})$ usando

as Eqs. B.26, B.27 e B.29 como

$$\begin{aligned}
Q(\mathbf{x}_1, \mathbf{x}_2) &= \begin{pmatrix} \mathbf{x}_1 - \boldsymbol{\mu}_1 & \mathbf{x}_2 - \boldsymbol{\mu}_2 \end{pmatrix}^T \begin{pmatrix} \boldsymbol{\Sigma}^{11} & \boldsymbol{\Sigma}^{12} \\ \boldsymbol{\Sigma}^{21} & \boldsymbol{\Sigma}^{22} \end{pmatrix} \begin{pmatrix} \mathbf{x}_1 - \boldsymbol{\mu}_1 \\ \mathbf{x}_2 - \boldsymbol{\mu}_2 \end{pmatrix} \\
&= (\mathbf{x}_1 - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}^{11} (\mathbf{x}_1 - \boldsymbol{\mu}_1) + (\mathbf{x}_1 - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}^{12} (\mathbf{x}_2 - \boldsymbol{\mu}_2) + \\
&\quad (\mathbf{x}_2 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{21} (\mathbf{x}_1 - \boldsymbol{\mu}_1) + (\mathbf{x}_2 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{22} (\mathbf{x}_2 - \boldsymbol{\mu}_2).
\end{aligned} \tag{B.30}$$

Explorando a simetria de $\boldsymbol{\Sigma}$, mais especificamente $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}^T$, e usando a propriedade comutativa do produto escalar juntamente da identidade matricial $\mathbf{B}^T \mathbf{A}^T = (\mathbf{A} \mathbf{B})^T$, encontramos que

$$\begin{aligned}
(\mathbf{x}_1 - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}^{12} (\mathbf{x}_2 - \boldsymbol{\mu}_2) &= (\mathbf{x}_1 - \boldsymbol{\mu}_1)^T (\boldsymbol{\Sigma}^{21})^T (\mathbf{x}_2 - \boldsymbol{\mu}_2) \\
&= (\boldsymbol{\Sigma}^{21} (\mathbf{x}_1 - \boldsymbol{\mu}_1))^T (\mathbf{x}_2 - \boldsymbol{\mu}_2) \\
&= (\mathbf{x}_2 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{21} (\mathbf{x}_1 - \boldsymbol{\mu}_1).
\end{aligned} \tag{B.31}$$

Assim, o expoente $Q(\mathbf{x}_1, \mathbf{x}_2)$ é reduzido a

$$\begin{aligned}
Q(\mathbf{x}_1, \mathbf{x}_2) &= (\mathbf{x}_1 - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}^{11} (\mathbf{x}_1 - \boldsymbol{\mu}_1) + 2(\mathbf{x}_2 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{21} (\mathbf{x}_1 - \boldsymbol{\mu}_1) + \\
&\quad (\mathbf{x}_2 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{22} (\mathbf{x}_2 - \boldsymbol{\mu}_2).
\end{aligned} \tag{B.32}$$

Entretanto, precisamos representar Q em termos dos blocos da matriz de covariância, não de sua inversa, para explicitar a forma gaussiana da distribuição marginal. Para tanto, passamos a relacionar os respectivos blocos dessas matrizes. Vamos utilizar a seguinte identidade para a inversa de uma matriz escrita em blocos [72]:

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}^{-1} = \begin{pmatrix} (\mathbf{A} - \mathbf{B} \mathbf{D}^{-1} \mathbf{C})^{-1} & -\mathbf{A}^{-1} \mathbf{B} (\mathbf{D} + \mathbf{C} \mathbf{D}^{-1} \mathbf{B})^{-1} \\ -\mathbf{D}^{-1} \mathbf{C} (\mathbf{A} - \mathbf{B} \mathbf{D}^{-1} \mathbf{C})^{-1} & (\mathbf{D} - \mathbf{C} \mathbf{A}^{-1} \mathbf{B})^{-1} \end{pmatrix}. \tag{B.33}$$

Essa identidade nos permite escrever os blocos de $\boldsymbol{\Sigma}^{-1}$ a partir dos blocos de $\boldsymbol{\Sigma}$ comparando diretamente as Eqs. B.29 e B.33, ou seja,

$$\begin{aligned}
\begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}^{-1} &= \begin{pmatrix} (\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21})^{-1} & -\boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12} (\boldsymbol{\Sigma}_{22} + \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{12})^{-1} \\ -\boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} (\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21})^{-1} & (\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12})^{-1} \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{\Sigma}^{11} & \boldsymbol{\Sigma}^{12} \\ \boldsymbol{\Sigma}^{21} & \boldsymbol{\Sigma}^{22} \end{pmatrix}.
\end{aligned} \tag{B.34}$$

Analisando a última igualdade, notamos que

$$\Sigma^{11} = (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}, \quad (\text{B.35})$$

$$\Sigma^{21} = -\Sigma_{22}^{-1}\Sigma_{21}(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}, \quad (\text{B.36})$$

$$\Sigma^{22} = (\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1}. \quad (\text{B.37})$$

Essas expressões para os blocos de Σ^{-1} introduzem considerável complexidade à Eq. B.32. Por isso, antes de as substituírmos em Q , usamos a identidade de Woodbury² [72] para reescrever Σ^{22} como

$$\Sigma^{22} = \Sigma_{22}^{-1} + \Sigma_{22}^{-1}\Sigma_{21}(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}\Sigma_{12}\Sigma_{22}^{-1}, \quad (\text{B.38})$$

de modo que os blocos Σ^{11} , Σ^{21} e Σ^{22} tenham um termo em comum.

Coletando esses resultados, podemos reescrever Q (Eq. B.32) como

$$\begin{aligned} Q(\mathbf{x}_1, \mathbf{x}_2) &= (\mathbf{x}_1 - \boldsymbol{\mu}_1)^T [(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}] (\mathbf{x}_1 - \boldsymbol{\mu}_1) - \\ &\quad 2(\mathbf{x}_2 - \boldsymbol{\mu}_2)^T [-\Sigma_{22}^{-1}\Sigma_{21}(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}] (\mathbf{x}_1 - \boldsymbol{\mu}_1) + \\ &\quad (\mathbf{x}_2 - \boldsymbol{\mu}_2)^T [\Sigma_{22}^{-1} + \Sigma_{22}^{-1}\Sigma_{21}(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}\Sigma_{12}\Sigma_{22}^{-1}] (\mathbf{x}_2 - \boldsymbol{\mu}_2) \\ &= (\mathbf{x}_2 - \boldsymbol{\mu}_2)^T \Sigma_{22}^{-1} (\mathbf{x}_2 - \boldsymbol{\mu}_2) + \\ &\quad (\mathbf{x}_2 - \boldsymbol{\mu}_2)^T [\Sigma_{22}^{-1}\Sigma_{21}(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}\Sigma_{12}\Sigma_{22}^{-1}] (\mathbf{x}_2 - \boldsymbol{\mu}_2) - \\ &\quad 2(\mathbf{x}_2 - \boldsymbol{\mu}_2)^T [\Sigma_{22}^{-1}\Sigma_{21}(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}] (\mathbf{x}_1 - \boldsymbol{\mu}_1) + \\ &\quad (\mathbf{x}_1 - \boldsymbol{\mu}_1)^T (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} (\mathbf{x}_1 - \boldsymbol{\mu}_1). \end{aligned} \quad (\text{B.39})$$

Escrevendo $\mathbf{u} = \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)$, $\mathbf{v} = (\mathbf{x}_1 - \boldsymbol{\mu}_1)$ e $\mathbf{A} = (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}$, temos

$$Q(\mathbf{x}_1, \mathbf{x}_2) = (\mathbf{x}_2 - \boldsymbol{\mu}_2)^T \Sigma_{22}^{-1} (\mathbf{x}_2 - \boldsymbol{\mu}_2) + \mathbf{u}^T \mathbf{A} \mathbf{u} - 2\mathbf{u}^T \mathbf{A} \mathbf{v} + \mathbf{v}^T \mathbf{A} \mathbf{v} \quad (\text{B.40})$$

ou, ainda,

$$Q(\mathbf{x}_1, \mathbf{x}_2) = (\mathbf{x}_2 - \boldsymbol{\mu}_2)^T \Sigma_{22}^{-1} (\mathbf{x}_2 - \boldsymbol{\mu}_2) + (\mathbf{v} - \mathbf{u})^T \mathbf{A} (\mathbf{v} - \mathbf{u}), \quad (\text{B.41})$$

² $(\mathbf{A} + \mathbf{C}\mathbf{B}\mathbf{D})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{C}(\mathbf{B}^{-1} + \mathbf{D}\mathbf{A}^{-1}\mathbf{C})^{-1}\mathbf{D}\mathbf{A}^{-1}$.

assumindo que a matriz \mathbf{A} é simétrica³. Mais explicitamente,

$$Q(\mathbf{x}_1, \mathbf{x}_2) = (\mathbf{x}_2 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}_{22}^{-1} (\mathbf{x}_2 - \boldsymbol{\mu}_2) + \left[(\mathbf{x}_1 - \boldsymbol{\mu}_1) - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} (\mathbf{x}_2 - \boldsymbol{\mu}_2) \right]^T \mathbf{A} \left[(\mathbf{x}_1 - \boldsymbol{\mu}_1) - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} (\mathbf{x}_2 - \boldsymbol{\mu}_2) \right]. \quad (\text{B.42})$$

Para tornar essa expressão do expoente Q mais similar à sua forma original, definimos

$$\mathbf{m} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} (\mathbf{x}_2 - \boldsymbol{\mu}_2) \quad (\text{B.43})$$

e

$$\mathbf{S} = (\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21}) \quad (\text{B.44})$$

para escrever

$$\mathbf{Q}(\mathbf{x}_1, \mathbf{x}_2) = (\mathbf{x}_1 - \mathbf{m})^T \mathbf{S}^{-1} (\mathbf{x}_1 - \mathbf{m}) + (\mathbf{x}_2 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}_{22}^{-1} (\mathbf{x}_2 - \boldsymbol{\mu}_2). \quad (\text{B.45})$$

A distribuição normal multivariada (Eq. B.25) pode ser reescrita como

$$f(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{(2\pi)^{\frac{p}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (\mathbf{x}_1 - \mathbf{m})^T \mathbf{S}^{-1} (\mathbf{x}_1 - \mathbf{m}) \right\} \times \exp \left\{ -\frac{1}{2} (\mathbf{x}_2 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}_{22}^{-1} (\mathbf{x}_2 - \boldsymbol{\mu}_2) \right\} \quad (\text{B.46})$$

³Sejam \mathbf{u} e \mathbf{v} vetores coluna e \mathbf{A} uma matriz simétrica ($\mathbf{A} = \mathbf{A}^T$). Logo,

$$\begin{aligned} \mathbf{u}^T \mathbf{A} \mathbf{u} - 2\mathbf{u}^T \mathbf{A} \mathbf{v} + \mathbf{v}^T \mathbf{A} \mathbf{v} &= \mathbf{u}^T \mathbf{A} \mathbf{u} - \mathbf{u}^T \mathbf{A} \mathbf{v} - \mathbf{u}^T \mathbf{A} \mathbf{v} + \mathbf{v}^T \mathbf{A} \mathbf{v} \\ &= \mathbf{u}^T \mathbf{A} (\mathbf{u} - \mathbf{v}) - (\mathbf{u}^T - \mathbf{v}^T) \mathbf{A} \mathbf{v} \\ &= \mathbf{u}^T \mathbf{A} (\mathbf{u} - \mathbf{v}) - (\mathbf{u} - \mathbf{v})^T \mathbf{A} \mathbf{v} \\ &= \mathbf{u}^T \mathbf{A} (\mathbf{u} - \mathbf{v}) - (\mathbf{u} - \mathbf{v})^T \mathbf{A} \mathbf{v} \\ &= \mathbf{u}^T \mathbf{A} (\mathbf{u} - \mathbf{v}) - (\mathbf{u} - \mathbf{v})^T \mathbf{A}^T \mathbf{v} \\ &= \mathbf{u}^T \mathbf{A} (\mathbf{u} - \mathbf{v}) - (\mathbf{A} (\mathbf{u} - \mathbf{v}))^T \mathbf{v} \\ &= \mathbf{u}^T \mathbf{A} (\mathbf{u} - \mathbf{v}) - (\mathbf{A} (\mathbf{u} - \mathbf{v}))^T \mathbf{v} \\ &= \mathbf{u}^T \mathbf{A} (\mathbf{u} - \mathbf{v}) - \mathbf{v}^T \mathbf{A} (\mathbf{u} - \mathbf{v}) \\ &= (\mathbf{u}^T - \mathbf{v}^T) \mathbf{A} (\mathbf{u} - \mathbf{v}) \\ &= (\mathbf{u} - \mathbf{v})^T \mathbf{A} (\mathbf{u} - \mathbf{v}) \\ &= (\mathbf{v} - \mathbf{u})^T \mathbf{A} (\mathbf{v} - \mathbf{u}). \end{aligned}$$

ou, expressando $|\Sigma|$ em função de seus blocos⁴ (Eq. B.28),

$$f(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{(2\pi)^{\frac{p}{2}}|\mathbf{S}|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2}(\mathbf{x}_1 - \mathbf{m})^T \mathbf{S}^{-1}(\mathbf{x}_1 - \mathbf{m}) \right\} \times \frac{1}{(2\pi)^{\frac{q}{2}}|\Sigma_{22}|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2}(\mathbf{x}_2 - \boldsymbol{\mu}_2)^T \Sigma_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2) \right\}. \quad (\text{B.47})$$

Finalmente, usando a forma anterior de $f(\mathbf{x})$ podemos demonstrar que a distribuição marginal ou reduzida de $f(\mathbf{x}_2)$ é dada por

$$f(\mathbf{x}_2) = \int f(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_1 = \frac{1}{(2\pi)^{\frac{q}{2}}|\Sigma_{22}|^{\frac{1}{2}}} \exp \left\{ (\mathbf{x}_2 - \boldsymbol{\mu}_2)^T \Sigma_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2) \right\} \quad (\text{B.48})$$

e que a distribuição condicional de \mathbf{x}_1 dado \mathbf{x}_2 é obtida fazendo

$$f(\mathbf{x}_1|\mathbf{x}_2) = \frac{f(\mathbf{x}_1, \mathbf{x}_2)}{f(\mathbf{x}_2)} = \frac{1}{(2\pi)^{\frac{p}{2}}|\mathbf{S}|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2}(\mathbf{x}_1 - \mathbf{m})^T \mathbf{S}^{-1}(\mathbf{x}_1 - \mathbf{m}) \right\}. \quad (\text{B.49})$$

Sendo assim, fica demonstrado que as distribuições marginal e condicional são normais multivariadas.

Retomando o propósito inicial de simular séries de movimento browniano fracionário, observamos que diversos algoritmos para a simulação desse processo estocástico podem ser encontrados na literatura [69]. O procedimento que escolhemos utilizar ao longo deste trabalho é um procedimento recursivo, conhecido como método de Hosking [52], o qual consiste em calcular sucessivamente a distribuição de probabilidade condicional das variáveis aleatórias de um processo gaussiano estacionário de média zero [69], como o ruído gaussiano fracionário. Amostras (séries temporais) de movimento browniano fracionário são obtidas da soma acumulada de amostras desse ruído.

Sejam as variáveis aleatórias $\{X_n, X_{n-1}, \dots, X_1, X_0\}$ um ruído gaussiano fracionário. A distribuição de probabilidade conjunta $f(\mathbf{x})$, que define o processo e dá origem a suas amostras, é da forma

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n+1}{2}}|\Gamma_n|^{\frac{1}{2}}} \exp \left\{ -\frac{\mathbf{x}^T \Gamma_n^{-1} \mathbf{x}}{2} \right\}, \quad (\text{B.50})$$

sendo $\mathbf{x} = (x_n, x_{n-1}, \dots, x_1, x_0)$ um vetor (coluna) de variáveis mudas e Γ_n a matriz de

⁴O determinante de uma matriz bloco \mathbf{M} pode ser escrito como [72]:

$$\det \mathbf{M} = \det \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} = \det(\mathbf{D}) \det(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}).$$

covariância. Explicitamente, essa matriz é escrita como

$$\mathbf{\Gamma}_n = \begin{pmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \cdots & \gamma(n) \\ \gamma(1) & \gamma(0) & \gamma(1) & \cdots & \gamma(n-1) \\ \gamma(2) & \gamma(1) & \gamma(0) & \cdots & \gamma(n-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma(n) & \gamma(n-1) & \gamma(n-2) & \cdots & \gamma(0) \end{pmatrix}_{(n+1) \times (n+1)} \quad (\text{B.51})$$

ou, ainda, $\mathbf{\Gamma}_n = (\gamma(|i-j|))_{i,j=0,\dots,n}$, de modo que cada elemento $\gamma(|i-j|)$ dessa matriz simétrica é calculado a partir da função de autocovariância do ruído gaussiano fracionário (Eq. B.22). Consequentemente, se estendermos a duração do processo, a distribuição conjunta das variáveis aleatórias $\{X_{n+1}, X_n, \dots, X_1, X_0\}$ é

$$f(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{n+2}{2}} |\mathbf{\Gamma}_{n+1}|^{\frac{1}{2}}} \exp \left\{ -\frac{\mathbf{y}^T \mathbf{\Gamma}_{n+1}^{-1} \mathbf{x}}{2} \right\}, \quad (\text{B.52})$$

sendo $\mathbf{y} = (x_{n+1}, x_n, \dots, x_1, x_0)$ um vetor de variáveis mudas e $\mathbf{\Gamma}_{n+1}$ a correspondente matriz de covariância. Desse modo,

$$\mathbf{y} = \begin{pmatrix} x_{n+1} \\ \mathbf{x} \end{pmatrix} \quad (\text{B.53})$$

e a distribuição condicional $f(x_{n+1}|\mathbf{x})$ é normal (comparando B.52 com B.25, B.48 e B.49). Queremos mostrar que essa distribuição é completamente determinada por \mathbf{x} e $\mathbf{\Gamma}_n$ e que, portanto, podemos estender o processo estocástico calculando recorrentemente essa distribuição.

Com esse objetivo, definimos o vetor (coluna) de covariância \mathbf{c}_n , cuja k -ésima componente é igual a $\gamma(k+1)$ (Eq. B.22), como

$$\mathbf{c}_n = \begin{pmatrix} \gamma(1) \\ \gamma(2) \\ \gamma(3) \\ \vdots \\ \gamma(n+1) \end{pmatrix}_{(n+1) \times 1}, \quad (\text{B.54})$$

sendo o vetor linha \mathbf{c}_n^T obtido da transposição de \mathbf{c}_n . Definimos também a matriz antidiagonal $\mathbf{F}_n = (\mathbf{1}(i=n-j))_{i,j=0,\dots,n}$ cuja função indicadora⁵ $\mathbf{1}(i=n-j)$ define seus elementos. A matriz \mathbf{F}_n , quando multiplicada pela direita por um vetor coluna, inverte esse vetor, de

⁵Uma função indicadora é definida como sendo igual a 1 quando a condição que define a função é satisfeita e 0 caso contrário. Especificamente, $\mathbf{1}(i=n-j)$ é igual a 1 quando o índice de linha i de um elemento de \mathbf{F}_n é igual à subtração do número n de linhas dessa matriz e de seu índice de coluna j . Caso essa condição não seja verificada, a função indicadora é igual a 0.

modo que a última componente do vetor se torna a primeira, a penúltima se torna a segunda e assim sucessivamente. Quando \mathbf{F}_n é multiplicada à esquerda por um vetor linha, uma inversão similar acontece. Explicitamente, temos

$$\mathbf{F}_n = \begin{pmatrix} \mathbf{1}(0 = n) & \cdots & \mathbf{1}(0 = n - (n - 1)) & \mathbf{1}(0 = n - n) \\ \mathbf{1}(1 = n) & \cdots & \mathbf{1}(1 = n - (n - 1)) & \mathbf{1}(1 = n - n) \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{1}(n - 1 = n) & \cdots & \mathbf{1}(n - 1 = n - (n - 1)) & \mathbf{1}(n - 1 = n - n) \\ \mathbf{1}(n = n) & \cdots & \mathbf{1}(1 = n - (n - 1)) & \mathbf{1}(n = n - n) \end{pmatrix}. \quad (\text{B.55})$$

Feitas essas definições, podemos escrever

$$\mathbf{\Gamma}_{n+1} = \begin{pmatrix} 1 & \mathbf{c}_n^T \\ \mathbf{c}_n & \mathbf{\Gamma}_n \end{pmatrix} \quad (\text{B.56})$$

$$= \begin{pmatrix} \mathbf{\Gamma}_n & \mathbf{F}_n \mathbf{c}_n \\ \mathbf{c}_n^T \mathbf{F}_n & 1 \end{pmatrix} \quad (\text{B.57})$$

e, definindo $\mathbf{d}_n = \mathbf{\Gamma}_n^{-1} \mathbf{c}_n$ e $\sigma_n^2 = 1 - \mathbf{c}_n^T \mathbf{\Gamma}_n^{-1} \mathbf{c}_n$, encontramos as seguintes expressões⁶ para a matriz de precisão:

$$\mathbf{\Gamma}_{n+1}^{-1} = \frac{1}{\sigma_n^2} \begin{pmatrix} 1 & -\mathbf{d}_n^T \\ -\mathbf{d}_n & \sigma_n^2 \mathbf{\Gamma}_n^{-1} + \mathbf{d}_n \mathbf{d}_n^T \end{pmatrix} \quad (\text{B.58})$$

$$= \frac{1}{\sigma_n^2} \begin{pmatrix} \sigma_n^2 \mathbf{\Gamma}_n^{-1} + \mathbf{F}_n \mathbf{d}_n \mathbf{d}_n^T \mathbf{F}_n & -\mathbf{F}_n \mathbf{d}_n \\ -\mathbf{d}_n^T \mathbf{F}_n & 1 \end{pmatrix}. \quad (\text{B.59})$$

Definindo ainda

$$\mu_n = \mathbf{c}_n^T \mathbf{\Gamma}_n^{-1} \begin{pmatrix} x_n \\ \vdots \\ x_1 \\ x_0 \end{pmatrix} = \mathbf{d}_n^T \begin{pmatrix} x_n \\ \vdots \\ x_1 \\ x_0 \end{pmatrix}, \quad (\text{B.60})$$

o expoente da distribuição de probabilidade conjunta do processo (Eq. B.52) pode ser reescrito como

$$\begin{pmatrix} x_{n+1} & \mathbf{x}^T \end{pmatrix} \mathbf{\Gamma}_{n+1}^{-1} \begin{pmatrix} x_{n+1} \\ \mathbf{x} \end{pmatrix} = \frac{(x_{n+1} - \mathbf{d}_n^T \mathbf{x})^2}{\sigma_n^2} + \mathbf{x}^T \mathbf{\Gamma}_n^{-1} \mathbf{x} \quad (\text{B.61})$$

$$= \frac{(x_{n+1} - \mu_n)^2}{\sigma_n^2} + \mathbf{x}^T \mathbf{\Gamma}_n^{-1} \mathbf{x}. \quad (\text{B.62})$$

⁶A primeira dessas igualdades segue diretamente da Eq. B.56 em conjunto com a identidade da Eq. B.33. A obtenção da segunda expressão, que segue da Eq. B.57, é consideravelmente mais complicada e passa pela utilização de outra identidade para a inversão em blocos de $\mathbf{\Gamma}_{n+1}$ e de condições de simetria e propriedades específicas das matrizes $\mathbf{\Gamma}_n$ e \mathbf{F}_n .

Logo, a distribuição de probabilidade conjunta (Eq. B.52) é dada por

$$f(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{n+2}{2}} |\mathbf{\Gamma}_{n+1}|^{\frac{1}{2}}} \exp \left\{ -\frac{\mathbf{x}^T \mathbf{\Gamma}_n^{-1} \mathbf{x}}{2} - \frac{1}{2} \frac{(x_{n+1} - \mu_n)^2}{\sigma_n^2} \right\}, \quad (\text{B.63})$$

e a distribuição condicional de x_{n+1} , dada uma realização particular do processo estocástico (comparando a Eq. B.63 com as Eqs. B.46 e B.49), é

$$f(x_{n+1}|\mathbf{x}) = \frac{1}{(2\pi\sigma_n^2)^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} \frac{(x_{n+1} - \mu_n)^2}{\sigma_n^2} \right\}. \quad (\text{B.64})$$

Essa distribuição é uma normal⁷ de média μ_n e variância σ_n^2 e fica completamente definida por \mathbf{x} e $\mathbf{\Gamma}_n$. Note, contudo, que é a matriz de precisão $\mathbf{\Gamma}_n^{-1}$ que surge no cálculo de μ_n e σ_n , o que implica em uma operação de inversão matricial a cada passo ao longo da geração de uma amostra. O procedimento de Hosking [52, 69] tem como mérito evitar essa operação de inversão. Mais precisamente, esse método consiste em usar as recorrências

$$\mathbf{d}_{n+1} = \begin{pmatrix} \mathbf{d}_n - \phi_n \mathbf{F}_n \mathbf{d}_n \\ \phi_n \end{pmatrix} \quad (\text{B.65})$$

e

$$\sigma_{n+1}^2 = \sigma_n^2 - \frac{(\gamma(n+2) - \tau_n)^2}{\sigma_n^2}, \quad (\text{B.66})$$

sendo definidos os escalares $\tau_n = \mathbf{d}_n^T \mathbf{F}_n \mathbf{c}_n = \mathbf{c}_n^T \mathbf{F}_n \mathbf{d}_n$ e

$$\phi_n = \frac{\gamma(n+2) - \tau_n}{\sigma_n^2} \quad (\text{B.67})$$

para calcular a média e variância necessárias para a determinação das distribuições condicionais de X_{n+2}, X_{n+3}, \dots evitando a operação de inversão da matriz de covariância. A Eq. B.65

⁷As definições anteriores de μ_n e σ_n não são arbitrárias ou, simplesmente, convenientes, mas decorrem das Eqs. B.43 e B.44 fazendo $\mu_n = \mathbf{m}$ (com $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = 0$) e $\sigma_n = \mathbf{S}$ (com $\boldsymbol{\Sigma} = \mathbf{\Gamma}_n$).

é encontrada usando a Eq. B.59 e as definições de \mathbf{c}_{n+1} e \mathbf{d}_{n+1} :

$$\begin{aligned}
\mathbf{d}_{n+1} &= \mathbf{\Gamma}_{n+1}^{-1} \mathbf{c}_{n+1} \\
&= \frac{1}{\sigma_n^2} \begin{pmatrix} \sigma_n^2 \mathbf{\Gamma}_n^{-1} + \mathbf{F}_n \mathbf{d}_n \mathbf{d}_n^T \mathbf{F}_n & -\mathbf{F}_n \mathbf{d}_n \\ -\mathbf{d}_n^T \mathbf{F}_n & 1 \end{pmatrix} \begin{pmatrix} \mathbf{c}_n \\ \gamma(n+2) \end{pmatrix} \\
&= \frac{1}{\sigma_n^2} \begin{pmatrix} \sigma_n^2 \mathbf{\Gamma}_n^{-1} \mathbf{c}_n + \mathbf{F}_n \mathbf{d}_n \mathbf{d}_n^T \mathbf{F}_n \mathbf{c}_n - \mathbf{F}_n \mathbf{d}_n \gamma(n+2) \\ -\mathbf{d}_n^T \mathbf{F}_n \mathbf{c}_n + \gamma(n+2) \end{pmatrix} \\
&= \frac{1}{\sigma_n^2} \begin{pmatrix} \sigma_n^2 \mathbf{d}_n + \mathbf{F}_n \mathbf{d}_n \tau_n - \mathbf{F}_n \mathbf{d}_n \gamma(n+2) \\ \gamma(n+2) - \tau_n \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{d}_n + \phi_n \mathbf{F}_n \mathbf{d}_n \\ \phi_n \end{pmatrix}.
\end{aligned} \tag{B.68}$$

A segunda recorrência, também decorre da Eq. B.59 e das definições de \mathbf{d}_{n+1} e \mathbf{c}_{n+1} :

$$\begin{aligned}
\sigma_{n+1}^2 &= 1 - \mathbf{c}_{n+1}^T \mathbf{d}_{n+1} \\
&= 1 - \frac{1}{\sigma_n^2} \begin{pmatrix} \mathbf{c}_n^T & \gamma(n+2) \end{pmatrix} \begin{pmatrix} \sigma_n^2 \mathbf{d}_n + \mathbf{F}_n \mathbf{d}_n \tau_n - \mathbf{F}_n \mathbf{d}_n \gamma(n+2) \\ \gamma(n+2) - \tau_n \end{pmatrix} \\
&= 1 - \frac{1}{\sigma_n^2} \left(\sigma_n^2 \mathbf{c}_n^T \mathbf{d}_n + \mathbf{c}_n^T \mathbf{F}_n \mathbf{d}_n \tau_n - \right. \\
&\quad \left. \mathbf{c}_n^T \mathbf{F}_n \mathbf{d}_n \gamma(n+2) - \gamma(n+2) \tau_n \right) \\
&= 1 - \frac{1}{\sigma_n^2} \left(\sigma_n^2 (1 - \sigma_n^2) + \tau_n^2 - \tau_n \gamma(n+2) - \gamma(n+2) \tau_n + \gamma^2(n+2) \right) \\
&= 1 - \frac{1}{\sigma_n^2} \left(\sigma_n^2 (1 - \sigma_n^2) + \tau_n^2 - 2\tau_n \gamma(n+2) + \gamma^2(n+2) \right) \\
&= 1 - (1 - \sigma_n^2) + \frac{(\gamma(n+2) - \tau_n)^2}{\sigma_n^2} \\
&= \sigma_n^2 + \frac{(\gamma(n+2) - \tau_n)^2}{\sigma_n^2}.
\end{aligned} \tag{B.69}$$

Sendo assim, a simulação de uma amostra de ruído gaussiano fracionário começa com o sorteio de um número aleatório x_0 obtido de uma distribuição normal de média zero e variância unitária. Em seguida, os parâmetros da distribuição condicional de X_1 são avaliados usando as recorrências anteriores, ou seja, $\mu_0 = \mathbf{d}_0 x_0 = \gamma(1)x_0$, $\sigma_0^2 = 1 - \mathbf{c}_0^T \mathbf{\Gamma}_0^{-1} \mathbf{c}_0 = 1 - \gamma^2(1)$ e $\tau_0 = \mathbf{d}_0^T \mathbf{F}_0 \mathbf{c}_0 = \gamma^2(1)$. Na sequência, outro número aleatório x_1 é sorteado da distribuição condicional $f(x_1|x_0)$. Repetindo esse procedimento para x_2, x_3, \dots obtemos uma amostra de ruído gaussiano fracionário $\{x_t\}_{t=0, \dots, n}$. Uma amostra de movimento browniano fracionário $\{y_t\}_{t=0, \dots, n}$ é obtida da soma acumulada do ruído, isto é, $y_t = \sum_{i=0}^t x_i$.

C.1 Aprendizagem estatística e algoritmo k -primeiros vizinhos

Algoritmos de aprendizagem estatística têm como objetivo extrair informação de conjuntos de dados [54, 73]. Considerando, por exemplo, as variáveis Y (variável dependente ou resposta) e X (variável independente ou preditora) e um conjunto de observações $\{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$, em geral, estamos interessados em estabelecer uma relação do tipo

$$\hat{y} = \hat{f}(X = x^*), \quad (\text{C.1})$$

na qual \hat{y} representa uma estimativa da variável Y obtida a partir de \hat{f} quando X assume o valor x^* . A estimativa da função \hat{f} é a essência dos métodos de aprendizagem estatística [54].

Uma distinção entre algoritmos de aprendizagem estatística pode ser estabelecida segundo diferenças no processo de aprendizagem. Por exemplo, quando os valores da variável dependente Y são conhecidos, a avaliação da precisão das predições é feita pela comparação direta entre os valores preditos e observados. Nesse caso, o processo de aprendizagem é dito supervisionado. O algoritmo de k -primeiros vizinhos, que infere a classe/valor da resposta de uma observação com base em um modelo construído usando observações prévias, constitui um exemplo de método supervisionado [54, 73]. Contrariamente, caso os valores de Y não sejam conhecidos, o processo de aprendizado é dito não supervisionado. Procedimentos de *clustering*, os quais tentam agrupar observações em categorias distintas a partir de suas características (preditores), constituem exemplos de algoritmos não supervisionados [54, 73].

Outra importante distinção entre esses métodos de aprendizagem estatística passa pelo

tipo da variável Y . Caso a variável Y assuma valores dentro de um conjunto finito de classes ou categorias possíveis, dizemos que a tarefa de previsão de Y é uma tarefa de classificação [54, 73]. Caso Y assuma valores numéricos contínuos, dizemos que a tarefa é de regressão [54, 73]. O método de k -primeiros vizinhos representa um exemplo conveniente de algoritmo que pode ser usado nesses dois tipos de tarefas [54, 73].

O algoritmo classificador de k -primeiros vizinhos prediz uma classe ou categoria \hat{y} para a resposta Y de uma observação $X = x^*$ usando um número k de observações vizinhas. Esses vizinhos são os pontos mais próximos (de acordo com alguma medida de distância, por exemplo, a distância euclidiana) em um espaço abstrato de mesma dimensão que a variável preditora X^1 [54]. O algoritmo infere uma distribuição de probabilidade a partir da vizinhança de uma observação x^* e prediz como resposta \hat{y} a classe mais provável dessa vizinhança [54, 73]. Posto de outra forma, esse algoritmo funciona segundo o “voto da maioria”, em que a classe mais popular entre os vizinhos de x^* é atribuída a \hat{y} (Figura C.1a). O desempenho do classificador de k -primeiros vizinhos pode ser medido pela fração de classificações corretas [54].

Quando usado em tarefas de regressão, o método de k -primeiros vizinhos atribui um valor numérico \hat{y} à resposta Y de uma observação $X = x^*$ seguindo a mesma estratégia de avaliar os valores de sua vizinhança. Denotando o conjunto dos k primeiros vizinhos de x^* por \mathcal{N}^* , o valor estimado para a resposta dessa observação é

$$\hat{y} = \hat{f}(x^*) = \frac{1}{k} \sum_{x_i \in \mathcal{N}^*} y_i, \quad (\text{C.2})$$

na qual x_i e y_i representam os valores observados para os preditores e resposta do i -ésimo vizinho de x^* . Sendo assim, é atribuído a \hat{y} o valor médio das respostas dos vizinhos de x^* (Figura C.1b). A avaliação das predições e, conseqüentemente, da qualidade do ajuste de k -primeiros vizinhos, pode ser feita usando o coeficiente de determinação R^2 definido pela expressão [54]

$$R^2 = 1 - \frac{SQ_{\text{tot}}}{SQ_{\text{res}}}, \quad (\text{C.3})$$

sendo $SQ_{\text{res}} = \sum_{i=1}^n (y_i - \hat{y}_i)$ a chamada soma residual de quadrados e $SQ_{\text{tot}} = \sum_{i=1}^n (y_i - \bar{y})$ a soma total de quadrados. A soma total de quadrados quantifica a variabilidade intrínseca ao conjunto de dados enquanto a soma residual de quadrados representa uma medida da quantidade de variabilidade que permanece não explicada após a regressão. Desse modo, o coeficiente de determinação mede a quantidade de variação da resposta que é explicada pelos preditores [54]. Esse coeficiente tem seu máximo igual a 1 quando a regressão explica a dependência entre Y e X perfeitamente e é igual a 0 caso a predição da variável independente

¹A variável preditora pode ser um vetor $X = (X_1, X_2, \dots, X_n)$ com n características preditivas da variável dependente Y . Desse modo, essas características formam um espaço n dimensional no qual a localização de uma observação é determinada pelas componentes de $x^* = (x_1^*, x_2^*, \dots, x_n^*)$.

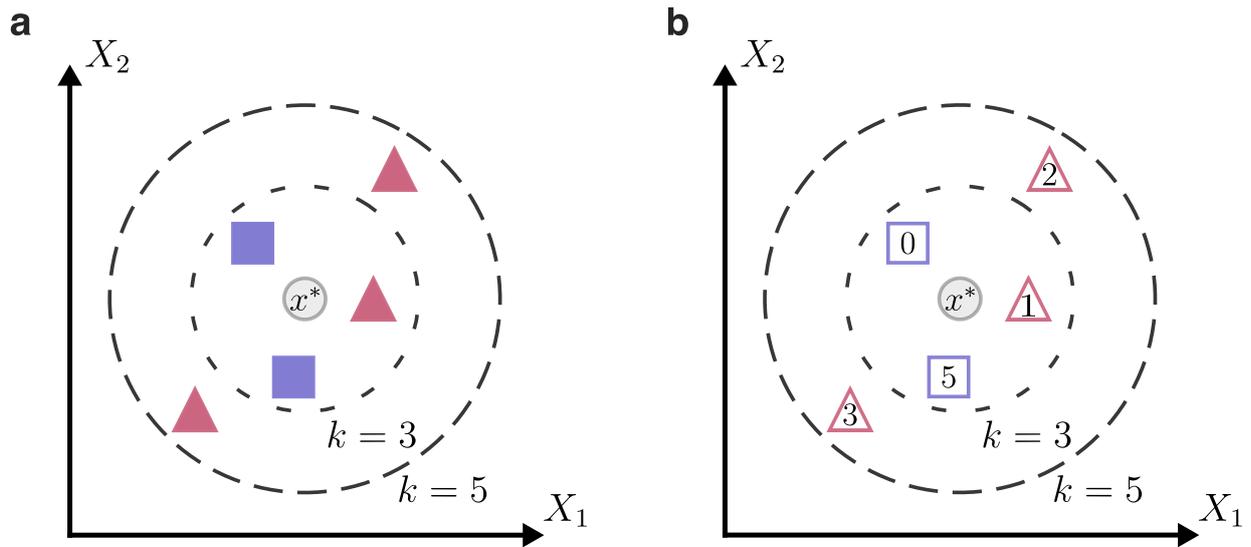


Figura C.1: Classificação e regressão de primeiros vizinhos. Uma observação $x^* = (x_1^*, x_2^*)$ e seus primeiros vizinhos são localizados em um espaço bidimensional associado à variável independente $X = (X_1, X_2)$. (a) Em uma tarefa de classificação, quando a resposta de x^* pertence a uma das classes no conjunto {"quadrado", "triângulo"}, a predição de primeiros vizinhos resulta em \hat{y} igual a "quadrado" quando $k = 3$ e igual a "triângulo" para $k = 5$. (b) Em uma tarefa de regressão, quando os vizinhos de x^* assumem valores numéricos (mostrados na figura), obtemos $\hat{y} = 2$ quando $k = 3$ e $\hat{y} = 2,2$ para $k = 5$.

seja sempre igual ao valor médio do conjunto de dados ($\hat{y}_i = \bar{y}$). O coeficiente de determinação também pode assumir valores negativos, indicando uma regressão arbitrariamente ruim.

C.2 Aplicação do algoritmo de k -primeiros vizinhos à estimativa do expoente de Hurst

A fim de prever o expoente de Hurst de séries de movimento browniano fracionário (Seção 2.3 do Capítulo 2), desenvolvemos uma tarefa de regressão e aplicamos o algoritmo de k -primeiros vizinhos [74]. Para tanto, usamos um *ensemble* de 4100 amostras desse processo estocástico, 100 amostras para cada $h \in \{0,10; 0,12; 0,14; \dots; 0,90\}$, e mapeamos todas essas séries temporais em redes complexas, mais especificamente, redes ordinais ($d = 2$) e grafos de quantil ($q = 50$). A partir dessas redes complexas, extraímos variáveis preditoras do expoente de Hurst: o caminho mais curto pesado médio ($\langle l \rangle$) das redes ordinais e o grau médio ($\langle k \rangle$) dos grafos de quantil. Desse modo, obtivemos os conjuntos de dados $\{(\langle l \rangle_1, h_1), \dots, (\langle l \rangle_{4100}, h_{4100})\}$ e $\{(\langle k \rangle_1, h_1), \dots, (\langle k \rangle_{4100}, h_{4100})\}$, mostrados na Figura C.2. Uma vez que o processo de aprendizagem do método de k -primeiros vizinhos é supervisão-

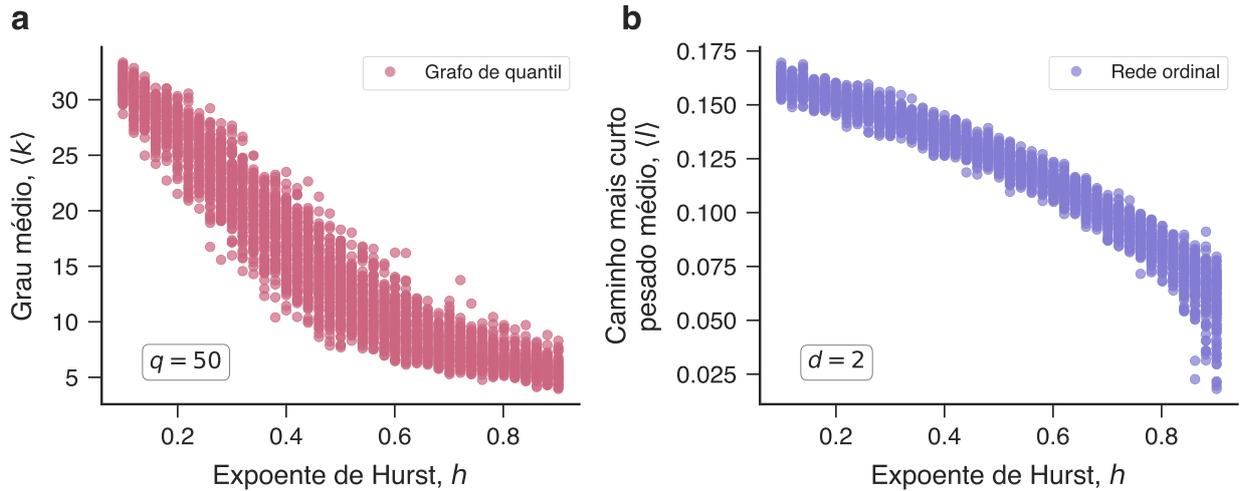


Figura C.2: Medidas de redes complexas e o expoente de Hurst. (a) Pares ordenados $\{(\langle k \rangle_1, h_1), \dots, (\langle k \rangle_{4100}, h_{4100})\}$ utilizados no treinamento dos modelos de k -primeiros vizinhos. Notamos que o grau médio $\langle k \rangle$ dos grafos de quantil diminui (em média) com o aumento do expoente de Hurst h . (b) Pares ordenados $\{(\langle l \rangle_1, h_1), \dots, (\langle l \rangle_{4100}, h_{4100})\}$ usados no treinamento de um modelo de k -primeiros vizinhos. O caminho mais curto pesado médio $\langle l \rangle$ das redes ordinais diminui (em média) com o aumento do expoente de Hurst h .

nado, separamos aleatoriamente 75% dos pares ordenados de cada um dos conjuntos para o procedimento de treinamento (conjuntos de treino) do algoritmo. Os 25% restantes foram separados para testar (conjuntos de teste) a precisão do modelo final extraído dos dados.

Para encontrar os melhores modelos de primeiros vizinhos para previsão do expoente de Hurst, procuramos pelo melhor número k de vizinhos a ser usado. Para tanto, avaliamos o desempenho do algoritmo para $k \in \{1, 2, 3, \dots, 499\}$ usando um procedimento de validação cruzada em cinco camadas [54, 73]. Nesse procedimento, cujos resultados são mostrados na Figura C.3, um conjunto de treino é dividido em cinco subconjuntos sendo quatro desses subconjuntos usados no processo de treinamento do modelo e o subconjunto restante (chamado de conjunto de validação) usado para avaliar o desempenho do modelo treinado. Repetido esse procedimento cinco vezes (de modo que cada subconjunto seja usado para validação), o valor ótimo de k é aquele cujo modelo treinado apresenta o melhor desempenho médio sobre os cinco conjuntos de validação. Para os dois modelos treinados, os valores de $k = 156$ (grafos de quantil) e $k = 94$ (redes ordinais) foram usados nos modelos finais. Em seguida, aplicamos os melhores modelos treinados aos seus respectivos conjuntos de teste. Os coeficientes de determinação obtidos foram $R^2 = 0,935$ (grafos de quantil) e $R^2 = 0,977$ (redes ordinais), indicando que o grau médio $\langle k \rangle$ dos grafos de quantil e o caminho mais curto médio pesado $\langle l \rangle$ das redes ordinais são bons preditores do expoente de Hurst. Além disso, a Figura C.3 mostra que o desempenho do algoritmo é bastante estável para um intervalo considerável de valores de k . Vale notar ainda que o desempenho do modelo final é sempre melhor (em

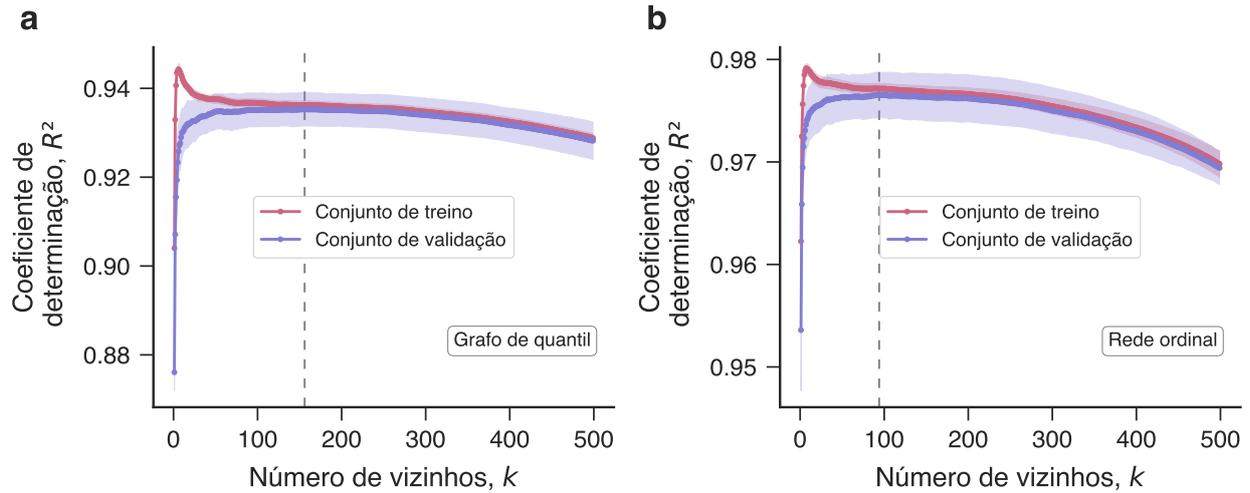


Figura C.3: Otimização do número de k -primeiros vizinhos. Coeficiente de determinação R^2 para modelos de k -primeiros vizinhos utilizando como variáveis predictoras do expoente de Hurst h (a) o grau médio dos grafos de quantil e (b) o caminho mais curto pesado médio das redes ordinais. As linhas sólidas representam o valor médio do coeficiente de determinação nos conjuntos de treino (em vermelho) e validação (em azul) e a região sombreada indica a banda de confiança de um desvio padrão. As linhas verticais tracejadas indicam os valores ótimos de k para cada um dos modelos.

média) no conjunto de treino do que no conjunto de validação e que a proximidade entre essas duas curvas é um bom indicativo de que o algoritmo escolhido é adequado para a tarefa de previsão de h a partir de medidas de redes complexas.

Análise de flutuação destendenciada

O método de análise de flutuação destendenciada (DFA — *Detrended Fluctuation Analysis*) é um procedimento muito utilizado para a estimativa de autocorrelações de longo alcance em séries temporais [55]. No caso de amostras de ruído gaussiano fracionário, em que a extensão das autocorrelações é controlada pelo expoente de Hurst, o método DFA pode ser usado na estimativa desse parâmetro [56].

De posse de uma série temporal $\{x_t\}_{t=1,\dots,N}$, o primeiro passo do método DFA consiste em obter a série integrada $\{y_t\}_{t=1,\dots,N}$, sendo

$$y_i = \sum_{j=1}^i x_j. \quad (\text{D.1})$$

Em seguida, dividimos a série integrada em partições (janelas) não sobrepostas contendo s elementos e ajustamos polinômios¹ a cada uma dessas partições a fim de remover tendências locais na série integrada. Tomando a diferença entre a série integrada e os polinômios ajustados obtemos uma série destendenciada $\{z_t\}_{t=1,\dots,N}$. Como próximo passo, calculamos a variância de cada um dos ν -ésimos segmentos da série destendenciada, isto é,

$$F_s^2(\nu) = \langle z_i^2 \rangle = \frac{1}{s} \sum_{i=1}^s z_{(\nu-1)s+i}^2. \quad (\text{D.2})$$

Finalmente, avaliamos o desvio padrão da variância dos segmentos destendenciados, também

¹Nos resultados apresentados neste trabalho, polinômios de primeiro grau foram ajustados às partições usando o método de mínimos quadrados. Além disso, o tamanho da maior partição usada é $\approx N/4$ [55].

chamado de função de flutuação [75] e definido como

$$F(s) = \left(\frac{1}{N/s} \sum_{\nu=1}^{N/s} F_s^2(\nu) \right)^{\frac{1}{2}}. \quad (\text{D.3})$$

Caso a série temporal $\{x_t\}_{t=1,\dots,N}$ seja correlacionada de longo alcance, a função de flutuação segue uma lei de escala da forma [75]

$$F(s) \sim s^\alpha. \quad (\text{D.4})$$

Para explicitar a relação entre o expoente α da função de flutuação, autocorrelações de longo alcance e o expoente de Hurst h , passamos a considerar uma série temporal $\{x_t\}_{t=1,\dots,N}$ estacionária e de média zero. Nesse caso, $z_i = y_i$ e temos

$$\begin{aligned} \langle y_i^2 \rangle &= \left\langle \left(\sum_{j=1}^i x_j \right) \left(\sum_{k=1}^i x_k \right) \right\rangle \\ &= \left\langle \sum_{j=1}^i x_j^2 \right\rangle + 2 \left\langle \sum_{\substack{j=1 \\ k>j}}^i x_j x_k \right\rangle \\ &= \sum_{j=1}^i \langle x_j^2 \rangle + 2 \sum_{\substack{j=1 \\ k>j}}^i \langle x_j x_k \rangle \\ &= i \langle x_j^2 \rangle + 2 \sum_{\substack{j=1 \\ k>j}}^i c(|k-j|) \\ &= i \langle x_j^2 \rangle + 2 \sum_{k=1}^{i-1} (i-k) c(k), \end{aligned} \quad (\text{D.5})$$

na qual $\langle x_j x_k \rangle = c(|k-j|)$ representa a autocovariância (Eq. B.5). Supondo $i \gg 1$ e que a função de autocovariância é do tipo $c(k) \sim k^{-\gamma}$, com $0 < \gamma < 1$, podemos usar as aproximações

$$\sum_{j=1}^{i-1} c(k) \sim \sum_{j=1}^i k^{-\gamma} \sim \int_1^i k^{-\gamma} dk \sim i^{1-\gamma} \quad (\text{D.6})$$

e

$$\sum_{j=1}^{1-i} k c(k) \sim \sum_{j=1}^i k^{1-\gamma} \sim \int_1^i k^{1-\gamma} dk \sim i^{2-\gamma}. \quad (\text{D.7})$$

Sendo assim, o comportamento assintótico de $\langle y_i^2 \rangle$ é da forma

$$\langle y_i^2 \rangle \sim i \langle x_j^2 \rangle + i^{-\gamma} + i^{2-\gamma} \sim i^{2-\gamma}, \quad (\text{D.8})$$

na qual o termo $i^{2-\gamma}$ é dominante para $0 < \gamma < 1$. Supondo que i faça o papel de s e substituindo na Eq. D.4, obtemos

$$F(s) \sim s^{1-\frac{\gamma}{2}}. \quad (\text{D.9})$$

Comparando a expressão anterior e a Eq. D.4, temos $\alpha = 1 - \frac{\gamma}{2}$. Se $\frac{1}{2} < h < 1$, o ruído gaussiano fracionário é correlacionado de longo alcance e $\gamma = 2 - 2h$ (Eq. B.24). Consequentemente, obtemos $\alpha = h$. Além disso, é possível mostrar que a relação $\alpha = h$ vale para todo $h \in (0, 1)$ [76]. Para o movimento browniano fracionário, um processo estocástico não estacionário, temos $\alpha = h + 1$ [77].

Referências Bibliográficas

- [1] Newman, M. *Networks: An Introduction* (Oxford University Press, 2010).
- [2] Dorogovtsev, S. *Lectures on Complex Networks*. Oxford Master Series in Physics (Oxford University Press, 2010).
- [3] Vespignani, A. Twenty years of network science. *Nature* **558**, 528–529 (2018).
- [4] Albert, R. & Barabási, A.-L. Statistical mechanics of complex networks. *Reviews of Modern Physics* 47–97 (2002).
- [5] Zou, Y., Donner, R. V., Marwan, N., Donges, J. F. & Kurths, J. Complex network approaches to nonlinear time series analysis. *Physics Reports* **787**, 1–97 (2019).
- [6] Gao, Z.-K., Small, M. & Kurths, J. Complex network analysis of time series. *EPL (Europhysics Letters)* **116**, 50001 (2016).
- [7] Lacasa, L., Luque, B., Ballesteros, F., Luque, J. & Nuño, J. C. From time series to complex networks: The visibility graph. *Proceedings of the National Academy of Sciences* **105**, 4972–4975 (2008).
- [8] Luque, B., Lacasa, L., Ballesteros, F. & Luque, J. Horizontal visibility graphs: Exact results for random time series. *Physical Review E* **80**, 046103 (2009).
- [9] Campanharo, A. S. L. O., Simer, M. I., Malmgren, R. D., Ramos, F. M. & Amaral, L. A. N. Duality between time series and networks. *PLoS ONE* **6**, e23378 (2011).
- [10] Small, M. Complex networks from time series: Capturing dynamics. In *2013 IEEE International Symposium on Circuits and Systems (ISCAS2013)*, 2509–2512 (2013).

- [11] Zhang, J. & Small, M. Complex network from pseudoperiodic time series: Topology versus dynamics. *Physical Review Letters* **96**, 238701 (2006).
- [12] Marwan, N., Donges, J. F., Zou, Y., Donner, R. V. & Kurths, J. Complex network approach for recurrence analysis of time series. *Physics Letters A* **373**, 4246–4254 (2009).
- [13] Mattmann, C. A. Computing: A vision for data science. *Nature* **493**, 473 (2013).
- [14] McCullough, M., Sakellariou, K., Stemler, T. & Small, M. Regenerating time series from ordinal networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **27**, 035814 (2017).
- [15] McCullough, M., Small, M., Stemler, T. & Iu, H. H.-C. Time lagged ordinal partition networks for capturing dynamics of continuous dynamical systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **25**, 053101 (2015).
- [16] Masoller, C., Hong, Y., Ayad, S., Gustave, F., Barland, S., Pons, A. J., Gómez, S. & Arenas, A. Quantifying sudden changes in dynamical systems using symbolic networks. *New Journal of Physics* **17**, 023068 (2015).
- [17] McCullough, M., Small, M., Iu, H. H. C. & Stemler, T. Multiscale ordinal network analysis of human cardiac dynamics. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **375**, 20160292 (2017).
- [18] Small, M., McCullough, M. & Sakellariou, K. Ordinal network measures — quantifying determinism in data. In *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, 1–5 (2018).
- [19] Pessa, A. A. B. & Ribeiro, H. V. Characterizing stochastic time series with ordinal networks. *Physical Review E* **100**, 042304 (2019).
- [20] Lacasa, L., Luque, B., Luque, J. & Nuño, J. C. The visibility graph: A new method for estimating the hurst exponent of fractional brownian motion. *EPL (Europhysics Letters)* **86**, 30001 (2009).
- [21] Bianchi, F. M., Livi, L., Alippi, C. & Jenssen, R. Multiplex visibility graphs to investigate recurrent neural network dynamics. *Scientific Reports* **7**, 44037 (2017).
- [22] Lacasa, L., Nuñez, A., Roldán, É., Parrondo, J. M. R. & Luque, B. Time series irreversibility: a visibility graph approach. *The European Physical Journal B* **85**, 217 (2012).

- [23] Bezsudnov, I. & Snarskii, A. From the time series to the complex networks: The parametric natural visibility graph. *Physica A: Statistical Mechanics and its Applications* **414**, 53–60 (2014).
- [24] Mandelbrot, B. B. & Van Ness, J. W. Fractional Brownian motions, fractional noises and applications. *SIAM Review* **10**, 422–437 (1968).
- [25] Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
- [26] Ding, M. & Yang, W. Distribution of the first return time in fractional brownian motion and its application to the study of on-off intermittency. *Physical Review E* **52**, 207–213 (1995).
- [27] Lacasa, L. & Toral, R. Description of stochastic and chaotic series using visibility graphs. *Physical Review E* **82**, 036120 (2010).
- [28] Rosso, O. A., Larrondo, H. A., Martin, M. T., Plastino, A. & Fuentes, M. A. Distinguishing noise from chaos. *Physical Review Letters* **99**, 154102 (2007).
- [29] Cencini, M., Falcioni, M., Olbrich, E., Kantz, H. & Vulpiani, A. Chaos or noise: Difficulties of a distinction. *Physical Review E* **62**, 427–437 (2000).
- [30] Zhang, R., Zou, Y., Zhou, J., Gao, Z.-K. & Guan, S. Visibility graph analysis for re-sampled time series from auto-regressive stochastic processes. *Communications in Nonlinear Science and Numerical Simulation* **42**, 396 – 403 (2017).
- [31] Ravetti, M. G., Carpi, L. C., Gonçalves, B. A., Frery, A. C. & Rosso, O. A. Distinguishing noise from chaos: objective versus subjective criteria using horizontal visibility graph. *PLoS ONE* **9**, e108004 (2014).
- [32] Watts, D. J. & Strogatz, S. H. Collective dynamics of ‘small-world’ networks. *Nature* **393**, 440–442 (1998).
- [33] Campanharo, A. S. L. O., Doescher, E. & Ramos, F. M. Application of quantile graphs to the automated analysis of EEG signals. *Neural Processing Letters* (2018).
- [34] Pineda, A. M., Ramos, F. M., Betting, L. E. & Campanharo, A. S. L. O. Use of complex networks for the automatic detection and the diagnosis of Alzheimer’s disease. In Rojas, I., Joya, G. & Catala, A. (eds.) *Advances in Computational Intelligence*, 115–126 (Springer, 2019).
- [35] Campanharo, A. S. & Ramos, F. M. Hurst exponent estimation of self-affine time series using quantile graphs. *Physica A: Statistical Mechanics and its Applications* **444**, 43–48 (2016).

- [36] Ott, E. *Chaos in Dynamical Systems* (Cambridge University Press, 2002).
- [37] Kantz, H. & Schreiber, T. *Nonlinear Time Series Analysis*. Cambridge Nonlinear Science Series (Cambridge University Press, 2004).
- [38] Eckmann, J.-P., Kamphorst, S. O. & Ruelle, D. Recurrence plots of dynamical systems. *Europhysics Letters (EPL)* **4**, 973–977 (1987).
- [39] Bradley, E. & Kantz, H. Nonlinear time-series analysis revisited. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **25**, 097610 (2015).
- [40] Packard, N. H., Crutchfield, J. P., Farmer, J. D. & Shaw, R. S. Geometry from a time series. *Physical Review Letters* **45**, 712–716 (1980).
- [41] May, R. M. Simple mathematical models with very complicated dynamics. *Nature* **261**, 459–467 (1976).
- [42] Strogatz, S. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering* (Avalon, 2014).
- [43] Bandt, C. & Pompe, B. Permutation entropy: A natural complexity measure for time series. *Phys. Rev. Lett.* **88**, 174102 (2002).
- [44] Cao, Y., Tung, W.-w., Gao, J. B., Protopopescu, V. A. & Hively, L. M. Detecting dynamical changes in time series using the permutation entropy. *Physical Review E* **70**, 046217 (2004).
- [45] Zunino, L., Soriano, M. C., Fischer, I., Rosso, O. A. & Mirasso, C. R. Permutation-information-theory approach to unveil delay dynamics from time-series analysis. *Physical Review E* **82**, 046212 (2010).
- [46] Unakafov, A. M. & Keller, K. Conditional entropy of ordinal patterns. *Physica D: Nonlinear Phenomena* **269**, 94–102 (2014).
- [47] Groth, A. Visualization of coupling in time series by order recurrence plots. *Physical Review E* **72**, 046220 (2005).
- [48] Sun, X., Small, M., Zhao, Y. & Xue, X. Characterizing system dynamics with a weighted and directed network constructed from time series data. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **24**, 024402 (2014).
- [49] Rössler, O. An equation for continuous chaos. *Physics Letters A* **57**, 397–398 (1976).
- [50] Jones, E., Oliphant, T., Peterson, P. *et al.* SciPy: Open source scientific tools for Python (2001–). URL <http://www.scipy.org/>.

- [51] Wills, P. & Meyer, F. G. Metrics for graph comparison: A practitioner’s guide. *arXiv preprint arXiv:1904.07414* (2019).
- [52] Hosking, J. R. M. Modeling persistence in hydrological time series using fractional differencing. *Water Resources Research* **20**, 1898–1908 (1984).
- [53] Gini, C. Measurement of inequality of incomes. *The Economic Journal* **31**, 124–126 (1921).
- [54] James, G., Witten, D., Hastie, T. & Tibshirani, R. *An Introduction to Statistical Learning: with Applications in R*. Springer Texts in Statistics (Springer New York, 2014).
- [55] Peng, C.-K., Buldyrev, S. V., Havlin, S., Simons, M., Stanley, H. E. & Goldberger, A. L. Mosaic organization of DNA nucleotides. *Physical Review E* **49**, 1685–1689 (1994).
- [56] Shao, Y.-H., Gu, G.-F., Jiang, Z.-Q., Zhou, W.-X. & Sornette, D. Comparing the performance of FA, DFA and DMA using different synthetic long-range correlated time series. *Scientific Reports* **2**, 835 (2012).
- [57] California Institute Of Technology And United States Geological Survey Pasadena - Southern California Seismic Network. Disponível: <http://www.fdsn.org/doi/10.7914/SN/CI>. Acessado: 28 Jun 2019.
- [58] Tectonic Summary – M 7.2 - 12km SW of Delta, B.C., MX. Disponível: <https://earthquake.usgs.gov/earthquakes/eventpage/usp000habu/executive>. Acessado: 28 Jun 2019.
- [59] Tectonic Summary – M 7.3 - Landers, California Earthquake. Disponível: <https://earthquake.usgs.gov/earthquakes/eventpage/usp00059sn/executive>. Acessado: 28 Jun 2019.
- [60] Tectonic Summary – M 7.1 - 16km SW of Ludlow, CA. Disponível: <https://earthquake.usgs.gov/earthquakes/eventpage/usp0009fwb/executive>. Acessado: 28 Jun 2019.
- [61] Utsu, T., Ogata, Y., S, R. & Matsu’ura. The centenary of the Omori formula for a decay law of aftershock activity. *Journal of Physics of the Earth* **43**, 1–33 (1995).
- [62] de Santana Costa, L. *Similaridades entre Terremotos e Ruídos Crepitantes em Carvão e Folhas Plásticas*. Tese de Doutorado, Universidade Estadual de Maringá (2017).
- [63] Erdős, P. & Rényi, A. On the evolution of random graphs. *Publications of Mathematical Institute of the Hungarian Academy of Science* **5**, 17–61 (1960).

- [64] Chatfield, C. *The Analysis of Time Series: An Introduction* (CRC Press, 2016).
- [65] Ross, S. M. *Introduction to Probability Models* (Elsevier, 2006).
- [66] Morettin, P. & de Castro Tolo, C. *Análise de Séries Temporais* (Edgard Blucher, 2006).
- [67] Mantegna, R. N. & Stanley, H. E. *Introduction to Econophysics: Correlations and Complexity in Finance* (Cambridge University Press, 1999).
- [68] Beran, J. *Statistics for Long-Memory Processes* (Taylor & Francis, 1994).
- [69] Diecker, T. *Simulation of fractional Brownian motion*. Ph.D. thesis, University of Twente (2004).
- [70] Rencher, A. & Schaalje, G. *Linear Models in Statistics* (Wiley, 2008).
- [71] Wang, R. Conditional and marginal of multivariate gaussian. Disponível: <http://fourier.eng.hmc.edu/e161/lectures/gaussianprocess/node5.html>. Acessado: 2 Set 2019.
- [72] Petersen, K. B. & Pedersen, M. S. The matrix cookbook. Disponível: <https://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf>. Acessado: 2 Set 2019.
- [73] Géron, A. *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems* (O'Reilly Media, 2017).
- [74] Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011).
- [75] Kantelhardt, J. W., Koscielny-Bunde, E., Rego, H. H., Havlin, S. & Bunde, A. Detecting long-range correlations with detrended fluctuation analysis. *Physica A: Statistical Mechanics and its Applications* **295**, 441–454 (2001).
- [76] Taqqu, M. S., Teverovsky, V. & Willinger, W. Estimators for long-range dependence: An empirical study. *Fractals* **03**, 785–798 (1995).
- [77] Peng, C., Havlin, S., Stanley, H. E. & Goldberger, A. L. Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **5**, 82–87 (1995).