



Ministério da
Ciência e Tecnologia



sid.inpe.br/mtc-m19/2011/03.10.23.03-TDI

DUALIDADE ENTRE A ANÁLISE DE SÉRIES TEMPORAIS E DE REDES COMPLEXAS

Andriana Susana Lopes de Oliveira Campanharo

Tese de Doutorado do Curso de Pós-Graduação em Computação Aplicada,
orientada pelo Dr. Fernando Manuel Ramos, aprovada em 29 de março de 2011

URL do documento original:

<<http://urlib.net/8JMKD3MGP7W/39AQPCB>>

INPE
São José dos Campos
2011

PUBLICADO POR :

Instituto Nacional de Pesquisas Espaciais - INPE

Gabinete do Diretor (GB)

Serviço de Informação e Documentação (SID)

Caixa Postal 515 - CEP 12.245-970

São José dos Campos - SP - Brasil

Tel.:(012) 3208-6923/6921

Fax: (012) 3208-6919

E-mail: pubtc@sid.inpe.br

CONSELHO DE EDITORAÇÃO E PRESERVAÇÃO DA PRODUÇÃO INTELLECTUAL DO INPE (RE/DIR-204):

Presidente:

Dr. Gerald Jean Francis Banon - Coordenação Observação da Terra (OBT)

Membros:

Dr^a Inez Staciarini Batista - Coordenação Ciências Espaciais e Atmosféricas (CEA)

Dr^a Maria do Carmo de Andrade Nono - Conselho de Pós-Graduação

Dr^a Regina Célia dos Santos Alvalá - Centro de Ciência do Sistema Terrestre (CST)

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Dr. Ralf Gielow - Centro de Previsão de Tempo e Estudos Climáticos (CPT)

Dr. Wilson Yamaguti - Coordenação Engenharia e Tecnologia Espacial (ETE)

Dr. Horácio Hideki Yanasse - Centro de Tecnologias Especiais (CTE)

BIBLIOTECA DIGITAL:

Dr. Gerald Jean Francis Banon - Coordenação de Observação da Terra (OBT)

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Yolanda Ribeiro da Silva Souza - Serviço de Informação e Documentação (SID)

EDITORAÇÃO ELETRÔNICA:

Vivéca Sant'Ana Lemos - Serviço de Informação e Documentação (SID)



Ministério da
Ciência e Tecnologia



sid.inpe.br/mtc-m19/2011/03.10.23.03-TDI

DUALIDADE ENTRE A ANÁLISE DE SÉRIES TEMPORAIS E DE REDES COMPLEXAS

Andriana Susana Lopes de Oliveira Campanharo

Tese de Doutorado do Curso de Pós-Graduação em Computação Aplicada,
orientada pelo Dr. Fernando Manuel Ramos, aprovada em 29 de março de 2011

URL do documento original:

<<http://urlib.net/8JMKD3MGP7W/39AQPCB>>

INPE
São José dos Campos
2011

Dados Internacionais de Catalogação na Publicação (CIP)

C15d Campanharo, Andriana Susana Lopes de Oliveira.
Dualidade entre a análise de séries temporais e de redes complexas / Andriana Susana Lopes de Oliveira Campanharo. – São José dos Campos : INPE, 2011.
xxvi+74 p. ; (sid.inpe.br/mtc-m19/2011/03.10.23.03-TDI)

Tese (Doutorado em Computação Aplicada) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2011.

Orientador : Dr. Fernando Manuel Ramos.

1. Sistemas complexos. 2. Redes complexas. 3. Análise de séries temporais. I. Título.

CDU 519.179.2

Copyright © 2011 do MCT/INPE. Nenhuma parte desta publicação pode ser reproduzida, armazenada em um sistema de recuperação, ou transmitida sob qualquer forma ou por qualquer meio, eletrônico, mecânico, fotográfico, reprográfico, de microfilmagem ou outros, sem a permissão escrita do INPE, com exceção de qualquer material fornecido especificamente com o propósito de ser entrado e executado num sistema computacional, para o uso exclusivo do leitor da obra.

Copyright © 2011 by MCT/INPE. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, microfilming, or otherwise, without written permission from INPE, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use of the reader of the work.

Aprovado (a) pela Banca Examinadora
em cumprimento ao requisito exigido para
obtenção do Título de Doutor(a) em
Computação Aplicada

Dr. José Demisio Simões da Silva



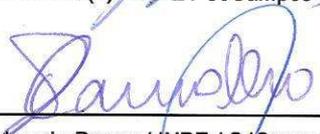
Presidente / INPE / SJCampos - SP

Dr. Fernando Manuel Ramos



Orientador(a) / INPE / SJCampos - SP

Dr. Solon Venâncio de Carvalho



Membro da Banca / INPE / SJCampos - SP

Dr. Haroldo Fraga de Campos Velho



Membro da Banca / INPE / São José dos Campos - SP

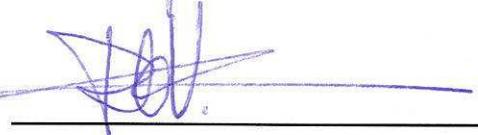
Dr. Joaquim José Barroso de Castro



Membro da Banca / INPE / SJCampos - SP

Dr. Luciano da Fontoura Costa

Dr. Ricardo Luiz Viana



Convidado(a) / UFPR / Curitiba - PR

Aluno (a): Andriana Susana Lopes de Oliveira Campanharo

São José dos Campos, 29 de março de 2011

“The mind that opens to a new idea never comes back to its original size”.

ALBERT EINSTEIN

A minha família

AGRADECIMENTOS

A Deus por sempre iluminar os meus caminhos e abençoar a minha vida. Também, por me conceder a sabedoria necessária para o desenvolvimento, conclusão e êxito deste trabalho.

A minha família por toda força, carinho e incentivo.

Ao meu orientador Dr. Fernando M. Ramos por todo apoio e orientação.

Ao meu orientador no exterior Dr. Luís A. N. Amaral por me conceder o privilégio de desenvolver dois anos de minha pesquisa de doutorado em seu laboratório. Agradeço também a todos os membros do “Amaral Lab”, em especial, aos colaboradores Dean Malmgren e Irmak Siner pelas valiosas discussões sobre o tema desta tese.

Ao Instituto Nacional de Pesquisas Espaciais - INPE pela oportunidade de prosseguir meus estudos e por todos os recursos físicos concedidos.

À Northwestern University pela oportunidade e recursos financeiros concedidos.

À Fundação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES pela bolsa de doutorado no Brasil. Ao Conselho de Desenvolvimento Científico e Tecnológico - CNPq pela bolsa de doutorado nos Estados Unidos.

À banca examinadora pelas valiosas sugestões e comentários visando o aprimoramento deste trabalho. Em especial, ao seu presidente Dr. José Demisio Simões da Silva, um professor querido, dedicado e entusiasta, que se foi e deixou saudades.

Aos professores do curso de Computação Aplicada - CAP por todo conhecimento compartilhado.

A todos os amigos que me apoiaram e me acompanharam durante este longo período.

A todos que contribuíram direta e indiretamente para que este trabalho fosse concretizado.

RESUMO

Recentemente, diversos mapeamentos de séries temporais em redes complexas foram propostos com o intuito de utilizar técnicas de caracterização de redes complexas para a análise de séries temporais. Embora esses mapeamentos demonstrem que séries temporais com dinâmicas distintas resultam em redes complexas com diferentes topologias, ainda não se sabe dizer como tais propriedades topológicas estão relacionadas com as séries originais. Nesta tese, propõe-se um mapeamento de uma série temporal em uma rede complexa com uma operação inversa aproximada, tornando possível a utilização de estatísticas em redes complexas para a caracterização de séries temporais e vice-versa. Como prova deste conceito, um conjunto de séries temporais é gerado, variando de periódica à aleatória e demonstra-se que o mapeamento proposto retém a maior parte da informação embutida na série temporal (ou rede) após a aplicação do mesmo e de seu inverso. Os resultados obtidos sugerem que a análise de redes complexas pode ser utilizada na distinção de regimes dinâmicos em séries temporais e, talvez o mais importante, a análise de séries temporais pode fornecer um conjunto de ferramentas úteis para a caracterização de redes complexas de uma forma totalmente original.

DUALITY BETWEEN TIME SERIES AND NETWORK ANALYSIS

ABSTRACT

Recently, several mappings from a time series to a network have been proposed with the intent of using network metrics to characterize time series. Although these mappings demonstrate that different time series result in networks with distinct topological properties, it remains unclear how these topological properties relate to the original time series. Here, it was proposed an intuitive mapping from a time series to a network with an approximate inverse operation, making it possible to use network statistics to characterize time series and time series statistics to characterize networks. As a proof of concept, an ensemble of time series ranging from periodic to random were generated and confirm that the proposed mapping retains much of the information encoded in the original time series and networks after repeated application of the mapping and its inverse. The results obtained in this thesis suggest that network analysis can be used to distinguish different dynamic regimes in time series and, perhaps more importantly, time series analysis can provide a powerful set of tools that augment the traditional network analysis toolkit to quantify networks in new and useful ways.

LISTA DE FIGURAS

	<u>Pág.</u>
2.1 O problema das pontes de Königsberg. (a) A cidade de Königsberg, atual Kaliningrado, Rússia com suas sete pontes conectando a ilha de Kneiphoff para as margens do rio Pregel. (b) Representação sistemática da área com as pontes. (c) Representação de Euler para o problema.	5
2.2 Exemplos de redes complexas: (a) Rede de contatos sexuais entre indivíduos; (b) Rede de contágios entre pessoas; (c) Rede dos amigos numa escola dos Estados Unidos; (d) Documentos num sítio da Web e ligações entre eles.	7
2.3 Representação gráfica de redes (a) não-direcionada, (b) direcionada e (c) com peso e não-direcionada com $N = 7$ vértices e $M = 14$ arestas cada. Na rede direcionada, as arestas com flechas indicam a direção com que os vértices são conectados. Na rede com peso, os valores de w_{ij} indicam os pesos das arestas, graficamente representados por arestas com diferentes espessuras.	8
2.4 Representação esquemática das relações entre os diferentes tipos de redes complexas.	9
2.5 Representação de redes complexas em matrizes de adjacência. (a) Rede não-direcionada: pares de vértices n_i e n_j , conectados por uma aresta, são representados pelo valor 1 na posição a_{ij} da matriz de adjacência A , caso contrário, pelo valor 0. (b) Rede direcionada: os elementos a_{ij} da matriz de adjacência A são iguais a 1 se existe uma aresta direcionada do vértice n_i para o vértice n_j	9
2.6 Exemplo de uma rede complexa com uma estrutura modular elevada. Note que os vértices em um mesmo grupo estão ligados por um número elevado de arestas, e que vértices em grupos diferentes estão ligados por um número reduzido de arestas.	14
2.7 Redes geradas a partir do modelo de Erdős e Rényi utilizando valores de p iguais a 0.0; 0.5 e 1.0, respectivamente. Com o incremento de p a rede torna-se cada vez mais conectada até que para $p = 1.0$ todos os vértices estão conectados uns aos outros.	16

2.8	Distribuição dos graus de conectividade de uma rede gerada a partir do modelo de Erdős e Rényi com 1,000 vértices, conectados com uma probabilidade $p = 0.2$	17
2.9	Procedimento de interpolação entre uma rede regular e uma rede aleatória, sem alterar o número de vértices ou arestas. Uma rede regular é obtida quando $p = 0.0$ e o valor $p = 1.0$ produz uma rede aleatória. Valores intermediários de p como, por exemplo, $p = 0.5$ produzem redes do tipo pequeno mundo.	18
2.10	Valores de C e L em função do parâmetro p	19
2.11	Distribuição do grau de conectividade para uma rede pequeno mundo formada por 1, 000 vértices, $k = 25$ vizinhos próximos e $p = 0.3$	20
2.12	Redes geradas a partir do modelo de Barabási e Albert com 20,30 e 50 vértices e $M = 1$, respectivamente. Em todos os casos pode-se observar a presença de um número reduzido de hubs.	21
2.13	Distribuição do grau de conectividade para uma rede livre de escala formada por 10,000 vértices e $M = 5$. Esta distribuição segue uma lei de potência da forma $P(k) \approx k^{-\gamma}$	22
2.14	Exemplos de séries temporais obtidas a partir da evolução das variáveis de estado $x(t)$, $y(t)$ e $z(t)$ em função do tempo t do Sistema (2.14), respectivamente.	25
2.15	Espectros de potências de séries temporais com dinâmicas distintas. (a) Série temporal aleatória uniformemente distribuída entre $[0, 1]$. (b) Série periódica (função seno). (c) Série temporal quasiperiódica $x(t)$, oriunda do sistema de Lorenz (2.14).	27
2.16	Funções de autocorrelação de séries temporais com dinâmicas distintas com $0 \leq \tau \leq 60$. (a) Ruído branco uniformemente distribuído entre $[0, 1]$. (b) Série periódica (função seno). (c) Série temporal $x(t)$ do sistema de Lorenz (2.14).	28
2.17	Séries temporais pseudo-periódicas (com mínimos locais representados por círculos, em vermelho) e suas representações em redes complexas. (a) A série temporal aleatória acrescida a uma taxa de ruído está associada a uma rede aleatória. (b) A série temporal caótica está associada a uma rede com características de redes pequeno mundo e livre de escala.	31

2.18	Exemplo de uma série temporal com 20 pontos e sua respectiva rede complexa, obtida a partir do algoritmo de visibilidade descrito anteriormente. Nesta rede cada vértice corresponde, na mesma ordem, a um ponto da série temporal. A visibilidade entre os pontos define a ligação entre os vértices na rede associada.	32
3.1	Ilustração do mapeamento proposto. <i>Mapeamento direto</i> : Uma série temporal X é mapeada em uma rede complexa g com $Q = 4$ vértices (quantis). Transições repetidas entre quantis resultam em arestas na rede com pesos maiores (representados por linhas mais grossas). <i>Mapeamento inverso</i> : A partir de uma rede complexa g , constituída de $Q = 4$ vértices, é construída uma série temporal X movendo-se aleatoriamente T passos de um vértice n_i para um vértice n_j , com probabilidade w_{ij}	36
4.1	Avaliação qualitativa da fidelidade do mapeamento proposto. Note que a primeira geração (1G) e a segunda geração (2G) de séries temporais e redes complexas possuem propriedades similares (Fig. 4.2-4.3), sugerindo-se que pode ser possível utilizar a análise de séries temporais para quantificar redes complexas, e vice-versa.	40
4.2	Comparação de algumas propriedades estatísticas associadas a primeira geração (1G) e a segunda geração (2G) de séries temporais. São comparadas as médias destas propriedades sob 10 realizações distintas. As barras de erro denotam os desvios padrão sob tais realizações. Para ambas as gerações de séries temporais, as funções de autocorrelação, os espectros de potência e as distribuições de seus valores são bastante similares, independente dos valores de p	41
4.3	Comparação de algumas propriedades topológicas associadas a primeira geração (1G) e a segunda geração (2G) de redes complexas. São comparados os valores médios de tais propriedades sob 10 realizações diferentes. As barras de erro denotam os desvios padrão sob tais realizações. Para ambas as gerações de redes complexas, as distribuições das forças de entrada, os pesos das arestas e os comprimentos dos menores caminhos são bastante similares, para todos os valores de p	43

4.4	Ilustração do mapeamento direto M_{QT} em séries temporais caóticas. São utilizados $T = 10,000$ pontos das variáveis $x(t)$ das equações de Lorenz e de Rossler e construídas redes complexas com $Q = 50$ quantis. As redes complexas resultantes apresentam diferenças nítidas em suas topologias. A rede associada ao sistema de Lorenz apresenta uma estrutura densa, onde os dois lóbulos do atrator são mapeados nos dois maiores módulos da rede complexa. Por outro lado, a rede complexa associada ao sistema de Rossler apresenta uma estrutura alongada decorrente do padrão quaseperiódico presente na série temporal correspondente.	45
4.5	Exemplos de séries temporais auto-similares com $-1.0 \leq \beta \leq 2.0$. Note que a medida que o valor de β aumenta as séries temporais correspondentes se tornam mais persistentes, ou seja, seus valores adjacentes se tornam mais correlacionados.	47
4.6	Espectros de potência (em escala log-log) associados as séries temporais auto-similares da Figura 4.5. Em todos os casos, os espectros de potência são caracterizados por uma lei de potência com inclinação dada por $-\beta$	48
4.7	Matrizes de adjacência com pesos obtidas através da aplicação do mapeamento direto M_{QT} nas séries temporais auto-similares da Figura 4.5. Quanto maior a anti-persistência na série temporal, maior é a probabilidade de vértices distantes estarem conectados. A medida que β aumenta, a correlação na série associada também aumenta e maior é a probabilidade de vértices vizinhos estarem conectados na rede correspondente.	49
4.8	Coeficiente de agrupamento e comprimento do caminho médio em função do coeficiente β associados as redes da Figura 4.7. São comparadas as médias destas propriedades sob 10 realizações distintas. As barras de erro denotam os desvios padrão sob tais realizações. Como esperado, um ruído branco ($\beta = 0.0$) gera uma rede completamente aleatória. Observe que séries com correlações crescentes (ou maior persistência) produzem redes cada vez mais regulares, onde o fenômeno de pequeno mundo tende a desaparecer.	50

4.9	Ilustração do mapeamento direto \mathcal{M}_{QT} no problema de detecção de diferenças na dinâmica das séries temporais associadas a pacientes em diferentes condições de saúde. São utilizadas séries temporais normalizadas de batimento cardíaco a 100-minutos de pacientes sadios e com falência cardíaca. Redes complexas são construídas utilizando-se $Q = 50$ quantis. As redes resultantes apresentam diferenças topológicas, com a presença de um cluster em separado na rede associada ao paciente não-sadio.	52
4.10	Três aplicações do mapeamento \mathcal{M}_{QT} utilizando-se $Q = 40, 60$ e 70 quantis em séries temporais de pacientes sadios e não-sadios. Note a similaridade entre tais redes com as redes apresentadas na Figura 4.9, atestando a robustez dos resultados associados ao mapeamento direto, independente dos valores de Q	53
4.11	Três aplicações do mapeamento \mathcal{M}_{QT} utilizando-se $Q = 50$ quantis em diferentes séries temporais associadas a pacientes sadios e não-sadios. Independente do número de pacientes, as redes complexas resultantes são visualmente similares as apresentadas na Figura 4.9.	54
4.12	Representações esquemáticas de duas redes complexas pertencentes a classes distintas. A primeira é a rede metabólica de <i>Arabidopsis thaliana</i> , que é caracterizada por uma estrutura em cadeias e um conjunto de hubs que são diretamente atingíveis entre si. A segunda é a rede de Internet nos Estados Unidos em 1997, que possui uma estrutura em estrela com diversos hubs.	55
4.13	Ilustração do mapeamento inverso em redes complexas reais de diferentes tipos — a rede metabólica <i>Arabidopsis thaliana</i> e a rede de Internet nos Estados Unidos em 1997. Primeiramente, vértices são associados a quantis através do reordenamento da matriz de adjacência com pesos correspondente (ambas em escala log-log). Em seguida, séries temporais são obtidas com $T = 100,000$ pontos cada utilizando-se redes com $Q = 607$ e $1,589$ vértices, respectivamente. As séries temporais resultantes apresentam diferenças nítidas em dinâmica.	56

4.14	Propriedades estatísticas das séries temporais apresentadas na Figura 4.13, obtidas da rede metabólica <i>Arabidopsis thaliana</i> e da rede de Internet nos Estados Unidos em 1997. As correlações de longo alcance presentes na rede metabólica são bem capturadas pela função de autocorrelação, pelo espectro de potência com inclinação $\beta = -1.4541$ e pelo valor médio do coeficiente de Hurst, dado por $H_u = 0.8579 \pm 0.0006$. Por outro lado, os resultados na rede de Internet mostram a assinatura de baixa correlação gerada por esta rede. Neste caso, o espectro do potência associado possui uma inclinação inferior, dada por $\beta = -1.1865$ e um coeficiente de Hurst também inferior, dado por $H_u = 0.7045 \pm 0.0014$	58
4.15	Quatro realizações do mapeamento inverso \mathcal{M}_{QT}^{-1} na rede metabólica <i>Arabidopsis thaliana</i> ($Q = 607$ e $T = 100,000$) e na rede de Internet nos Estados Unidos em 1997 ($Q = 1,589$ e $T = 100,000$). Note a similaridade destas séries com as séries apresentadas na Figura 4.13, demonstrando a robustez do mapeamento inverso proposto.	59
4.16	Aplicação do mapeamento inverso em redes complexas reais pertencentes a diferentes classes — a rede metabólica <i>Escherichia coli</i> , as redes de transporte aéreo da Ásia & Meio Leste e Norte Americana e a rede de iterações entre proteínas em <i>Saccharomyces cerevisia</i> . Primeiramente, as matrizes de adjacência com pesos associadas são reordenadas e a associação entre vértices e quantis é estabelecida. As séries temporais resultantes apresentam diferenças em dinâmica.	60

LISTA DE TABELAS

	<u>Pág.</u>
4.1 Características de diversas redes reais estudadas e das séries temporais correspondentes, obtidas através da aplicação do mapeamento inverso \mathcal{M}_{QT}^{-1}	61

LISTA DE SÍMBOLOS

\mathcal{M}_{QT}	–	mapeamento direto
\mathcal{M}_{QT}^{-1}	–	mapeamento inverso
g	–	rede complexa
N	–	número de vértices
M	–	número de arestas
A	–	matriz de adjacência sem pesos
W	–	matriz de adjacência com pesos
k	–	grau de um vértice
s	–	força de um vértice
$\langle k \rangle$	–	grau de conectividade médio
C	–	coeficiente de agrupamento médio
L	–	comprimento do menor caminho médio
E	–	eficiência global
M_o	–	modularidade
q	–	quantil
Q	–	número de quantis
X	–	série temporal
T	–	número de pontos de uma série temporal
A_{xx}	–	função de autocorrelação
C_{xy}	–	função de correlação cruzada
ρ	–	coeficiente de correlação linear
H_u	–	coeficiente de Hurst
β	–	coeficiente de persistência
τ	–	escala temporal
ω	–	frequência

SUMÁRIO

	<u>Pág.</u>
1 INTRODUÇÃO	1
2 MÉTODOS E APROXIMAÇÃO	5
2.1 Análise de redes complexas	5
2.1.1 Introdução histórica	5
2.1.2 Definições e conceitos básicos	7
2.1.3 Medidas estruturais em redes complexas	10
2.1.3.1 Grau de conectividade	10
2.1.3.2 Coeficiente de agrupamento	11
2.1.3.3 Comprimento do menor caminho	12
2.1.3.4 Modularidade	13
2.1.4 Modelos de redes complexas	14
2.1.4.1 O modelo de Erdős e Rényi (1959)	14
2.1.4.2 O modelo de Watts e Strogatz (1998)	15
2.1.4.3 O modelo de Barabási e Albert (1999)	19
2.2 Análise de séries temporais	23
2.2.1 Introdução	23
2.2.2 Técnicas de Análise de Séries Temporais	24
2.2.2.1 Análise espectral	24
2.2.2.2 Função de autocorrelação	26
2.2.2.3 Coeficiente de Hurst	29
2.2.3 Aproximação por Redes Complexas	29
2.2.3.1 O trabalho de Zhang e Small (2006)	29
2.2.3.2 O trabalho de Lacasa et al. (2008)	32
3 MAPEAMENTO PROPOSTO	35
4 RESULTADOS	39
4.1 Aplicação do mapeamento direto em séries temporais caóticas	44
4.2 Aplicação do mapeamento direto em séries temporais auto-similares	46
4.3 Aplicação do mapeamento direto em séries temporais de batimento cardíaco	51
4.4 Aplicação do mapeamento inverso em redes complexas reais	55

5	CONCLUSÕES	63
6	APÊNDICE A - O algoritmo de Floyd-Warshall	65
7	APÊNDICE B - Reordenamento da matriz de adjacência com pesos	67
	REFERÊNCIAS BIBLIOGRÁFICAS	69

1 INTRODUÇÃO

Redes complexas são descritas por um conjunto de vértices (nós) e arestas (conexões, ligações ou *links*) e algum tipo de interação entre os mesmos. Diversos sistemas de importância tecnológica ou mesmo social podem ser representados por redes complexas. Por exemplo, uma célula pode ser descrita como uma rede de componentes químicos conectados por reações químicas; a Internet é uma rede de roteadores e computadores conectados por conexões físicas ou sem fio; idéias se propagam através de uma rede social cujos vértices são os seres humanos e as arestas representam os diferentes tipos de relações sociais. Esses sistemas representam apenas alguns dos muitos exemplos que têm chamado a atenção da comunidade científica na investigação de seus mecanismos internos, sob a perspectiva em redes complexas (BARABÁSI, 2002).

Historicamente, o estudo de redes complexas tem sido de domínio de um ramo da matemática discreta conhecida como *Teoria de Grafos*. Desde o seu nascimento em 1736, quando o matemático suíço Leonardo Euler publicou a solução para o problema das sete pontes de Königsberg, esta teoria tem se desenvolvido e fornecido respostas para uma série de questões práticas, tais como colorir regiões de um mapa utilizando um número mínimo de cores ou ainda preencher n empregos por n pessoas com utilidade máxima (BOCCALETTI et al., 2006). Além de fazer uso de medidas desenvolvidas pela teoria dos grafos, a teoria de redes complexas utiliza conceitos provenientes da mecânica estatística, física não-linear e sistemas complexos.

De modo geral, as redes complexas são divididas em quatro diferentes categorias: redes sociais, redes de informação, redes tecnológicas e redes biológicas. Em uma rede social os vértices representam pessoas e as arestas representam interações entre as mesmas. Desta forma, a estrutura das relações existentes entre pessoas (LILJEROS et al., 2001) ou grupos de indivíduos de uma determinada especialidade, como artistas (GLEISER; DANON, 2003), cientistas (ALBERT et al., 2002; MATIA et al., 2005), colaboradores (GUIMERA et al., 2005) ou mesmo esportistas (ONODY; CASTRO, 2004), pode ser estudada sob uma perspectiva de redes complexas. Em uma rede de informação, os vértices representam algum tipo de informação (conhecimento) e as arestas representam relações entre essas informações. Um exemplo clássico de uma rede de informação é a rede de citações entre artigos acadêmicos. Neste caso, os vértices representam artigos e uma aresta direcionada de um artigo A para um artigo B indica que A cita B (REDNER, 1998). Um outro exemplo de uma rede de

informação é a WWW (World-Wide Web), uma rede de páginas de textos acessíveis através da Internet, na qual seus vértices são as páginas e suas arestas são hiperlinks (URLs) (HUBERMAN, 2001). As redes tecnológicas são sistemas onde o objeto de estudo são estruturas ou sistemas criados pelo ser humano para a distribuição de recursos. Exemplos de redes tecnológicas incluem a Internet (formada por roteadores conectados por fibras óticas) (FALOUTSOS et al., 1999) e a Teia Mundial (formada por portais, *blogs* e *sites* pessoais e educacionais que se referenciam) (ALBERT et al., 1999). No caso da Internet, fatores dinâmicos podem ser simulados, como propagação de falhas ou vírus entre roteadores. Por fim, redes complexas podem ser utilizadas para descrever vários sistemas biológicos. Um exemplo de rede biológica é a rede de interações entre proteínas, onde os vértices constituem proteínas e as arestas as interações entre as mesmas (MASLOV; SNEPPEN, 2002). As cadeias alimentares também podem ser representadas por redes formadas por espécies conectadas de acordo com relações de predatismo (AMARAL; MEYER, 1999; CAMACHO et al., 2007; STOUFFER et al., 2007; STOUFFER et al., 2006). Nestas redes, fenômenos dinâmicos podem ser simulados, a fim de quantificar os efeitos causados por desastres ambientais ou extinção de espécies (KRAUSE et al., 2003).

O interesse na área de redes complexas tem sido incrementado por diversos fatores. Primeiro, a informatização na aquisição de dados em diversas áreas tem levado ao surgimento de bases de dados amplas com informações detalhadas sobre a topologia de diversas redes reais. Segundo, o poder computacional disponível para os cientistas tem permitido a pesquisa de redes contendo milhões de vértices, levantando questões que não poderiam ser investigadas anteriormente. Terceiro, a proximidade cada vez maior entre diversas áreas da ciência tem permitido a investigação de propriedades comuns dos mais variados tipos de redes complexas. Estes três fatores combinados vêm permitindo à comunidade científica mover-se além das aproximações reducionistas do passado e investigar o comportamento das redes complexas como um todo (ALBERT; BARABÁSI, 2002).

Recentemente, a teoria de redes complexas foi utilizada em uma aplicação até então inédita: na análise da dinâmica de séries temporais. No contexto de sistemas dinâmicos, a análise de séries temporais é primariamente utilizada na identificação da natureza subjacente a um fenômeno de interesse através de uma sequência de observações. Atualmente, existe um grande número de técnicas de análise de séries temporais que varia desde métodos tempo-frequência, tais como as transformadas

de Fourier e Wavelet (KORNER, 1988; BOX et al., 2008; PERCIVAL; WALDEN, 2000), a métodos não-lineares, tais como reconstrução do espaço de fases, dimensão de correlação, expoentes de Lyapunov e entropias (STROGATZ, 1994; KANTZ; SCHREIBER, 2003; CAMPANHARO et al., 2008). Tais técnicas permitem pesquisadores sumarizar as características de uma série temporal, e desta forma, determinar a dinâmica subjacente de um sistema ou prever como o mesmo evolui com o tempo.

Obviamente, estas medidas não preservam necessariamente *todas* as propriedades de uma série temporal, e desta forma, novas técnicas têm sido propostas para capturar informações adicionais ou quantificar séries temporais de novas formas (ZHANG et al., 2006; LAI et al., 2010; VERPLANCKE et al., 2010; AO, 2010). Um dos avanços mais interessantes foi sugerido pelo trabalho de Zhang e Small (2006) e outros (LACASA et al., 2008; LUQUE et al., 2009; YANG; YANG, 2008; MARWAN et al., 2009; XU. et al., 2008; GAO; JIN, 2009), que propuseram diferentes mapeamentos de séries temporais em redes complexas baseados em conceitos como visibilidade (LACASA et al., 2008; LUQUE et al., 2009), correlações (YANG; YANG, 2008), análise de recorrência (MARWAN et al., 2009) e reconstrução do espaço de fase (XU. et al., 2008; GAO; JIN, 2009). Estes estudos demonstraram que, independentemente do mapeamento utilizado, séries temporais com características distintas possuem representações em redes complexas com topologias distintas. Este fato sugere que pode ser possível diferenciar propriedades de uma série temporal utilizando medidas em redes complexas, contudo não é ainda claro como a análise de séries temporais pode ser útil na investigação das propriedades topológicas de uma rede complexa.

Na raiz deste problema está o fato que a maior parte destes mapeamentos $\mathcal{M} : \mathcal{T} \rightarrow \mathcal{G}$ de uma série temporal $X \in \mathcal{T}$ em uma rede complexa $g \in \mathcal{G}$ (ZHANG; SMALL, 2006; LACASA et al., 2008; LUQUE et al., 2009; YANG; YANG, 2008; MARWAN et al., 2009; XU. et al., 2008; GAO; JIN, 2009) não possui uma operação inversa $\mathcal{M}^{-1} : \mathcal{G} \rightarrow \mathcal{T}$. Se tal mapeamento e seu inverso existissem, seria possível criar uma representação dual entre uma série temporal e uma rede complexa e relacionar diretamente propriedades estatísticas entre as mesmas. Não somente a análise de séries temporais se beneficiaria da pesquisa relacionada a redes complexas (NEWMAN, 2003; COSTA et al., 2007), mas a teoria de redes complexas seria capaz de utilizar mais de três séculos de pesquisas teóricas e aplicadas em séries temporais. Neste trabalho, cumpre-se este objetivo introduzindo um mapeamento inédito de uma série temporal em uma rede complexa que possui uma operação inversa aproximada.

Os capítulos restantes desta tese estão organizados da seguinte maneira:

- Capítulo 2: Neste capítulo é feita uma descrição das principais medidas topológicas disponíveis na literatura para a caracterização de redes complexas. É feita também uma breve descrição de algumas ferramentas utilizadas na análise de séries temporais;
- Capítulo 3: Neste capítulo é apresentado um mapeamento inédito de uma série temporal em uma rede complexa (mapeamento direto), como também, sua operação inversa aproximada (mapeamento inverso). São também discutidas as principais características dos mesmos;
- Capítulo 4: Neste capítulo são apresentados os principais resultados obtidos a partir da aplicação do mapeamento direto e de seu inverso em um conjunto de séries temporais e redes complexas reais e sintéticas, respectivamente;
- Capítulo 5: Com base nas análises realizadas no Capítulo 4, neste capítulo são apresentadas as conclusões obtidas, como também algumas sugestões para trabalhos futuros.

2 MÉTODOS E APROXIMAÇÃO

2.1 Análise de redes complexas

2.1.1 Introdução histórica

Historicamente, o estudo de redes complexas tem sido de domínio de um ramo da matemática discreta conhecida como *Teoria de Grafos*. Esta teoria iniciou-se com o trabalho de Leonhard Euler para resolver o famoso problema das *Sete Pontes de Königsberg*. Como afirmado por Euler em seu manuscrito: “Na cidade de Königsberg na Prússia há uma ilha *A*, chamada Kneiphoff com dois braços do rio Pregel circulando ao redor dela. Existem também setes pontes *a, b, c, d, e, f* e *g* cortando os braços do rio Pregel. A questão é se uma pessoa pode caminhar de tal forma a atravessar cada uma dessas pontes uma e não mais que uma vez” (conforme Figura 2.1).

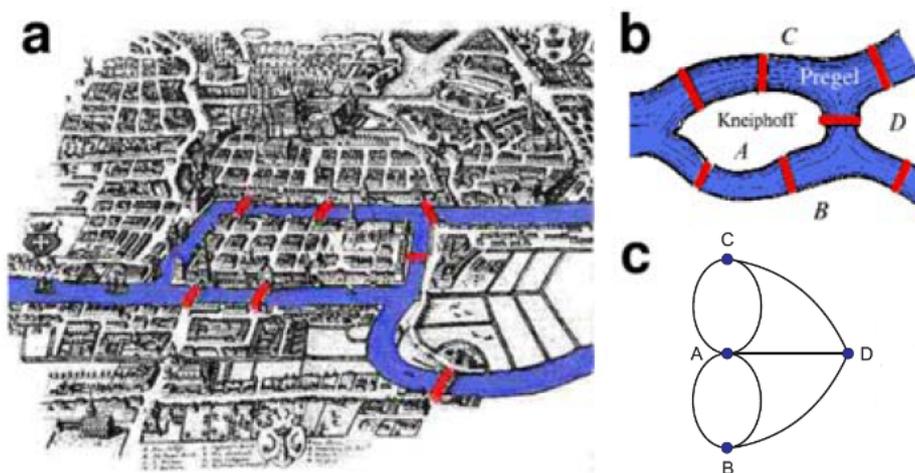


Figura 2.1 - O problema das pontes de Königsberg. (a) A cidade de Königsberg, atual Kaliningrado, Rússia com suas sete pontes conectando a ilha de Kneiphoff para as margens do rio Pregel. (b) Representação sistemática da área com as pontes. (c) Representação de Euler para o problema.

Fonte: (AMARAL; OTTINO, 2004)

Euler representou a topologia da região em um grafo, onde regiões de terra correspondiam vértices e pontes suas arestas criando, possivelmente, o primeiro grafo da história. Para a resolução deste problema, Euler dividiu os vértices em graus pares e ímpares e então demonstrou que:

- A soma dos graus dos vértices de um grafo é par;
- Todo grafo tem um número par de vértices com grau ímpar.

Estes resultados permitiram Euler mostrar que:

- Se o número de vértices com grau ímpar é maior que dois então não existe caminho Euleriano — um caminho Euleriano é um caminho entre dois vértices arbitrários no qual todas as arestas de um grafo aparecem exatamente uma vez;
- Se o número de vértices com grau ímpar é exatamente dois, existe um caminho Euleriano a partir de um destes vértices com grau ímpar;
- Sem a presença de vértices com grau ímpar, o caminho Euleriano pode ser iniciado a partir de qualquer vértice arbitrário.

Portanto, como todos os quatro vértices do problema das pontes de Königsberg eram ímpares, Euler demonstrou que não existia solução para este problema, ou seja, não existia nenhum caminho que permitisse atravessar cada ponte uma única vez.

Desde seu nascimento em 1736, a teoria de grafos forneceu respostas para questões práticas como por exemplo: distribuir tarefas entre pessoas com a máxima eficiência; colorir regiões de um mapa usando um número mínimo de cores ou ainda preencher n empregos por n pessoas com utilidade máxima (BOCCALETTI et al., 2006). A teoria das redes complexas nasceu da aplicação de medidas desenvolvidas pela teoria dos grafos e conceitos provenientes da mecânica estatística, física não-linear e sistemas complexos (COSTA et al., 2007). Redes complexas são descritas por um conjunto de vértices (nós) e arestas (conexões, ligações ou *links*) e algum tipo de interação entre os mesmos (conforme Figura 2.2).

Embora existam semelhanças, a teoria das redes complexas difere da teoria dos grafos em três aspectos básicos: (i) ela está relacionada com a modelagem de redes reais, por meio de análise de dados empíricos; (ii) em geral, as redes estudadas não são estáticas, mas evoluem no tempo, modificando sua estrutura; (iii) as redes, muitas vezes, não são consideradas apenas objetos topológicos, mas constituem estruturas onde processos dinâmicos (como a propagação de doenças e opiniões) podem ser simulados (COSTA et al., 2007).

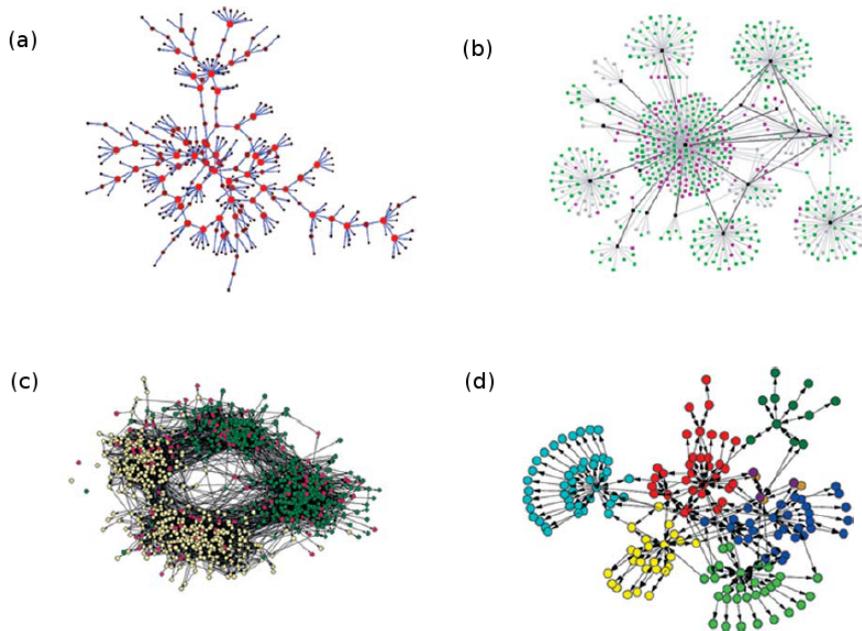


Figura 2.2 - Exemplos de redes complexas: (a) Rede de contatos sexuais entre indivíduos; (b) Rede de contágios entre pessoas; (c) Rede dos amigos numa escola dos Estados Unidos; (d) Documentos num sítio da Web e ligações entre eles. Fonte: (MENDES, 2005)

2.1.2 Definições e conceitos básicos

Matematicamente, uma rede $g = (\mathcal{N}, \mathcal{L})$ é formada por um conjunto de N vértices, $\mathcal{N} = \{n_1, n_2, \dots, n_N\}$ e um conjunto de M arestas, $\mathcal{L} = \{l_1, l_2, \dots, l_M\}$. As arestas presentes podem ser direcionadas - quando o sentido da ligação é considerado, e não-direcionadas - quando o sentido da ligação não é considerado. Se as arestas possuem intensidade, a cada aresta é associado um peso, e neste caso, a rede é representada por $g = (\mathcal{N}, \mathcal{L}, \mathcal{W})$, onde $\mathcal{W} = \{w_1, w_2, \dots, w_M\}$ constitui um conjunto de pesos das M arestas. Exemplos de redes não-direcionadas, direcionadas e com peso são apresentadas na Figura 2.3.

Em uma rede complexa não-direcionada e sem peso, cada uma de suas arestas é definida por um par de vértices n_i e n_j , e denotada como (n_i, n_j) ou l_{ij} . Dois vértices n_i e n_j conectados por uma aresta são referenciados como *adjacentes*. Quando um par de vértices não-adjacentes n_i e n_j pode ser conectado por uma sequência de m arestas distintas $(n_i, l_1), (l_1, l_2), \dots, (l_{m-1}, n_j)$, diz-se que este conjunto de arestas constitui um *caminho* entre os vértices n_i e n_j e que m é o *comprimento do caminho*

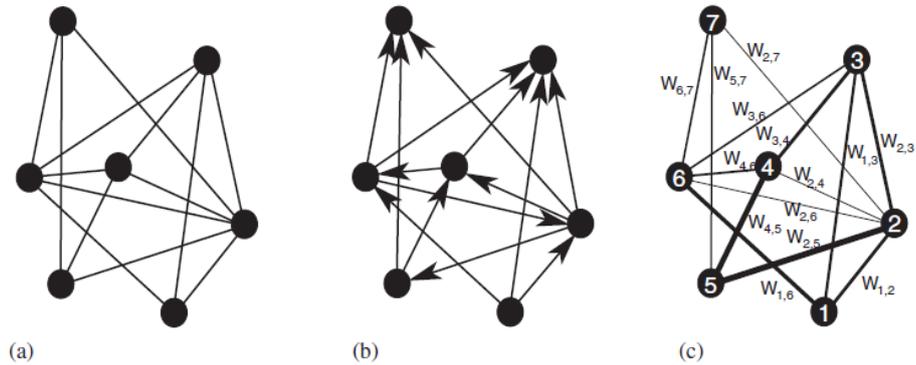


Figura 2.3 - Representação gráfica de redes (a) não-direcionada, (b) direcionada e (c) com peso e não-direcionada com $N = 7$ vértices e $M = 14$ arestas cada. Na rede direcionada, as arestas com flechas indicam a direção com que os vértices são conectados. Na rede com peso, os valores de w_{ij} indicam os pesos das arestas, graficamente representados por arestas com diferentes espessuras.

Fonte: (BOCCALETTI et al., 2006)

entre os mesmos.

As redes complexas podem ser *estáticas*, quando não há variação do número de vértices, arestas ou na configuração das mesmas; ou *dinâmicas*, quando há a variação de pelo menos uma destas componentes. O tipo mais comum de rede é aquele cujas arestas são direcionadas e com peso (COSTA et al., 2007). A partir deste tipo de rede mais geral é possível obter as demais configurações através de operações de *limiarização*, para a obtenção de redes sem peso, e *simetrização*, para a obtenção de redes não-direcionadas. A limiarização é realizada retirando-se as arestas cujo peso é menor do que um limiar pré-definido e associando pesos unitários as arestas remanescentes. Já a simetrização transforma as arestas direcionadas em não-direcionadas.

Em termos computacionais, uma rede complexa pode ser armazenada em dois tipos diferentes de estruturas de dados: matrizes de adjacência ou listas. Na representação em matrizes de adjacência, se dois vértices n_i e n_j estão conectados, a posição a_{ij} da matriz de adjacência A será igual a um, caso contrário igual a zero. Em uma rede com peso a posição w_{ij} da matriz de adjacência com pesos W será igual ao peso da aresta entre n_i e n_j . Na representação em listas, apenas os pares de vértices (n_i, n_j) conectados por uma aresta são armazenados. Para redes com peso, a lista

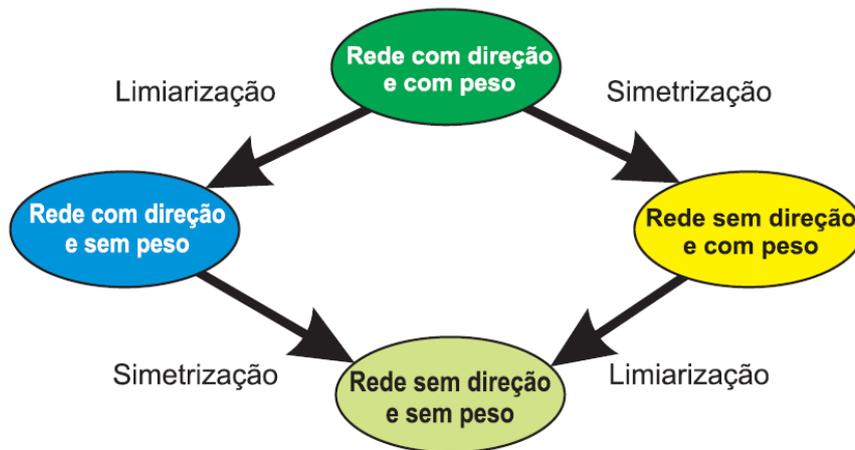


Figura 2.4 - Representação esquemática das relações entre os diferentes tipos de redes complexas.

Fonte: (RODRIGUES, 2007)

possui um terceiro elemento relacionado ao peso das aresta, (n_i, n_j, w_{ij}) . A Figura 2.5 mostra um exemplo de armazenamento de uma rede não-direcionada e de uma rede direcionada em matrizes de adjacências.

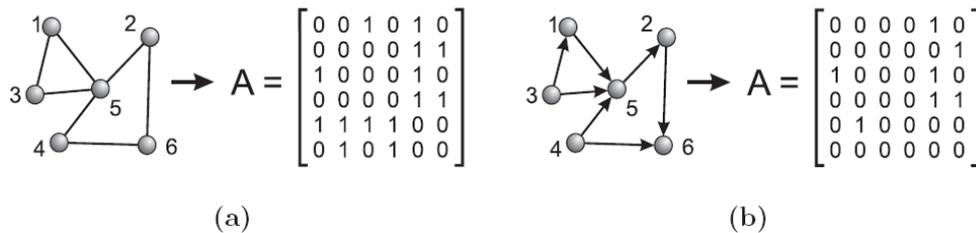


Figura 2.5 - Representação de redes complexas em matrizes de adjacência. (a) Rede não-direcionada: pares de vértices n_i e n_j , conectados por uma aresta, são representados pelo valor 1 na posição a_{ij} da matriz de adjacência A , caso contrário, pelo valor 0. (b) Rede direcionada: os elementos a_{ij} da matriz de adjacência A são iguais a 1 se existe uma aresta direcionada do vértice n_i para o vértice n_j .

Cada estrutura de armazenamento tem suas vantagens e desvantagens. O uso de listas permite maior economia de memória, principalmente se a rede é do tipo esparsa. Contudo, o acesso as ligações é mais complexo devido a necessidade de buscas. O

acesso as ligações em matrizes de adjacências é mais simples, porém, este tipo de estrutura de dado é somente bem aproveitado em redes densas (ou seja, redes que possuem a maior parte de seus vértices conectados entre si) (RUGGIERO et al., 2011).

2.1.3 Medidas estruturais em redes complexas

A pesquisa em redes complexas iniciou-se com o esforço em definir-se medidas para caracterizar a topologia de redes reais. O resultado principal tem sido a identificação de uma série de propriedades estatísticas comuns para grande parte das redes reais consideradas (NEWMAN, 2003; ALBERT; BARABÁSI, 2002). Alguns dos conceitos mais utilizados no estudo de redes complexas serão discutidos a seguir, e ainda, tais medidas serão aplicadas nas redes construídas no Capítulo 4.

2.1.3.1 Grau de conectividade

Por definição, o grau de conectividade de um vértice n_i , em uma rede não-direcionada com N vértices, é o número de suas arestas incidentes a outros vértices, ou seja:

$$k_i = \sum_j a_{ij} = \sum_j a_{ji}, \quad (2.1)$$

onde a_{ij} ($i = 1, 2, \dots, N$ e $j = 1, 2, \dots, N$) são elementos da matriz de adjacência A associada. O grau de conectividade médio é definido como a média aritmética dos graus de conectividade sobre todos os vértices, ou seja:

$$\langle k \rangle = \frac{1}{N} \sum_i k_i = \frac{1}{N} \sum_{ij} a_{ij}. \quad (2.2)$$

No caso de redes direcionadas, cada um de seus vértices possui dois tipos de graus de conectividade: o grau de conectividade de entrada (número total de arestas de entrada) e o grau de conectividade de saída (número total de arestas de saída):

$$k_i^{\text{entrada}} = \sum_j a_{ji}, \quad (2.3)$$

$$k_i^{\text{saída}} = \sum_j a_{ij}. \quad (2.4)$$

Note que, no caso de redes direcionadas, o grau total de conectividade de um vértice n_i é dado por $k_i = k_i^{\text{entrada}} + k_i^{\text{saída}}$. Os graus de conectividade de entrada e saída médios são os mesmos, ou seja:

$$\langle k^{\text{out}} \rangle = \langle k^{\text{in}} \rangle = \frac{1}{N} \sum_{ij} a_{ij}. \quad (2.5)$$

Finalmente, no caso de redes com peso para cada vértice é associada uma quantidade denominada força (em inglês, strength), definida como a soma do peso de suas arestas:

$$s_i^{\text{entrada}} = \sum_j w_{ji}, \quad (2.6)$$

$$s_i^{\text{saída}} = \sum_j w_{ij}, \quad (2.7)$$

onde w_{ij} ($i = 1, 2, \dots, N$ e $j = 1, 2, \dots, N$) são elementos da matriz de adjacência com pesos W associada.

A caracterização topológica mais básica de uma rede g pode ser obtida em termos da distribuição $P(k)$, definida como a probabilidade que um vértice escolhido aleatoriamente tenha grau de conectividade k . Em redes reais, a distribuição $P(k)$ desvia-se significativamente de uma distribuição de Poisson¹ esperada para uma rede aleatória e, em muitos casos, exibe característica de uma lei de potência com um expoente γ entre 2 e 3 (BOCCALETTI et al., 2006). As redes com estas características são chamadas de redes “livre de escala” (em inglês, scale free).

2.1.3.2 Coeficiente de agrupamento

Também conhecido como transitividade, o agrupamento é uma propriedade típica de redes de conhecimento, onde dois indivíduos com um amigo em comum são também conhecidos um do outro (WASSERMAN; FAUST, 1994). Esta tendência inerente ao agrupamento é quantificada pelo coeficiente de agrupamento (WATTS; STROGATZ, 1998). Dada uma rede não-direcionada g , suponha que um vértice n_i tenha k_i vizi-

¹A distribuição de Poisson é um tipo de distribuição discreta de probabilidade, cuja forma analítica é dada por $P(k) = \frac{\lambda^k e^{-\lambda}}{k!}$, onde λ é o parâmetro da distribuição.

nhos, logo $k_i(k_i - 1)/2$ arestas podem existir entre tais vizinhos (isto ocorre quando todos os vizinhos de n_i estão conectados entre si). A razão entre o número E_i de arestas que realmente existem pelo número total de arestas possíveis fornece o valor do coeficiente de agrupamento C_i do vértice n_i , ou seja:

$$C_i = \frac{2E_i}{k_i(k_i - 1)} . \quad (2.8)$$

Define-se o coeficiente de agrupamento médio da rede como a média dos C_i para todos os vértices n_i ($i = 1, 2, \dots, N$). Pela Equação (2.8) pode-se observar que o coeficiente de agrupamento assume valores pertencentes ao intervalo $[0, 1]$. As definições do coeficiente de agrupamento para redes direcionadas e/ou com peso podem ser encontradas em [Fagiolo \(2007\)](#).

2.1.3.3 Comprimento do menor caminho

Em uma rede sem peso e não-direcionada, o comprimento do menor caminho entre dois vértices n_i e n_j é um caminho que os conecta cujo comprimento é mínimo, ou seja, um caminho com o menor número possível de arestas. No caso de redes com peso, os pesos das arestas podem ser interpretados de duas maneiras distintas. Primeiramente, eles podem estar relacionados com alguma distância física, como por exemplo, em uma rede onde os vértices constituem cidades e as arestas as distâncias entre as mesmas. Neste caso, a distância ao longo de um caminho pode ser considerada como a soma dos pesos das arestas naquele caminho. Segundo, os pesos podem refletir a força de conexão entre os vértices como, por exemplo, em uma rede de Internet onde os vértices constituem roteadores e os pesos das arestas as larguras de banda. Neste caso, a distância ao longo de um caminho pode ser considerada como a soma dos inversos dos pesos das arestas naquele caminho ([COSTA et al., 2007](#)).

Os menores caminhos entre todos os pares de vértices de uma rede podem ser representados através de uma matriz de distâncias D , cujos elementos d_{ij} representam o valor do menor caminho entre os vértices n_i e n_j . O menor caminho médio L de uma rede conectada é então obtido com base na média dos valores da matriz D , ou seja:

$$L = \frac{1}{N(N-1)} \sum_{i \neq j} d_{ij}. \quad (2.9)$$

Um problema encontrado com a definição (2.9) é que o valor de L diverge se existir pelo menos uma componente desconectada na rede associada. Uma possibilidade para evitar tal divergência é limitar a soma apenas a pares de vértices pertencentes a maior componente conectada da rede (WATTS; STROGATZ, 1998). Outra possibilidade é utilizar a definição de *eficiência global* (LATORA; MARCHIORI, 2001):

$$E = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{1}{d_{ij}}, \quad (2.10)$$

que quantifica a eficiência da rede no envio de informação entre vértices, assumindo que a mesma é proporcional ao inverso da distância entre dois vértices n_i e n_j .

Existem vários métodos numéricos que calculam o menor caminho entre quaisquer pares de vértices de uma rede complexa, tais como os algoritmos de Dijkstra, Floyd-Warshall, Bellman-Ford, etc. Neste trabalho, o valor de L foi obtido com base no algoritmo de Floyd-Warshall e a descrição do mesmo pode ser encontrada no Capítulo 6.

2.1.3.4 Modularidade

Uma grande parte das redes reais, incluindo redes sociais, biológicas e tecnológicas, possui uma estrutura modular. Ou seja, arestas densamente distribuídas entre vértices pertencentes a um mesmo grupo e esparsamente distribuídas entre vértices pertencentes a grupos distintos (Fig. 2.6). A modularidade é uma medida que quantifica a divisão de uma rede em módulos. Para uma dada partição de vértices em uma rede em módulos, a modularidade \mathcal{M}_o desta partição é definida como:

$$\mathcal{M}_o \equiv \sum_{s=1}^r \left[\frac{l_s}{L} - \left(\frac{d_s}{2L} \right)^2 \right], \quad (2.11)$$

onde r é o número de módulos, L é o número de arestas na rede, l_s é o número de arestas entre vértices no módulo s , e d_s é a soma dos graus dos vértices no módulo s . Esta definição de modularidade implica que $\mathcal{M}_o \leq 1$ e que $\mathcal{M}_o = 0$ em uma partição aleatória de vértices. A modularidade M_o de uma rede é definida como a maior modularidade de todas as possíveis partições $M_o = \max \mathcal{M}_o$ (GUIMERÁ et al., 2004; SALES-PARDO et al., 2007).

Além das medidas apresentadas anteriormente, diversas medidas têm sido propostas

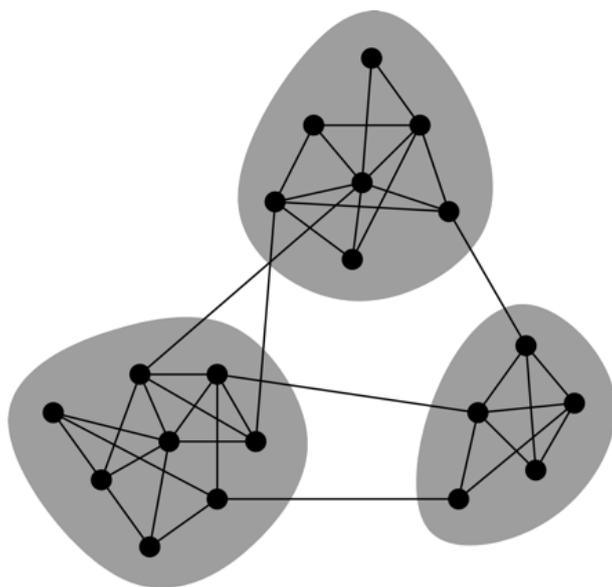


Figura 2.6 - Exemplo de uma rede complexa com uma estrutura modular elevada. Note que os vértices em um mesmo grupo estão ligados por um número elevado de arestas, e que vértices em grupos diferentes estão ligados por um número reduzido de arestas.

Fonte: (NEWMAN, 2006)

na literatura com a finalidade de se analisar a estrutura e o funcionamento de uma rede complexa. As descrições de tais medidas podem ser encontradas nos trabalhos de Costa et al. (2007) e de Barabási (2002).

2.1.4 Modelos de redes complexas

2.1.4.1 O modelo de Erdős e Rényi (1959)

Em 1959, dois matemáticos húngaros, Paul Erdős e Alfred Rényi criaram um modelo de rede baseado em ligações aleatórias, que ficou conhecido como *grafos aleatórios de Erdős e Rényi* (ERDŐS; RÉNYI, 1959). De acordo com este modelo, uma rede aleatória é construída a partir de um conjunto de N vértices totalmente desconectados e a cada passo dois vértices são escolhidos aleatoriamente e conectados com uma probabilidade fixa p , sendo que cada par de vértices é considerado uma única vez. Redes complexas geradas por este modelo possuem uma estrutura altamente homogênea.

A Figura 2.7 mostra as redes obtidas a partir do processo descrito acima, para

três diferentes realizações. Pode-se observar que para $p = 0.0$, a rede é totalmente desconectada, com o incremento de p a rede torna-se cada vez mais conectada até que para $p = 1.0$ todos os vértices estão conectados uns aos outros.

A Figura 2.8 apresenta um histograma da distribuição do grau de conectividade para uma rede aleatória com 1,000 vértices. Pode-se observar que este histograma segue uma distribuição de Poisson, o que significa que, em média, a maior parte dos vértices tem o mesmo grau de conectividade e que poucos vértices possuem graus de conectividade inferior ou superior à média. Além disso, as redes geradas por este modelo possuem valores de L pequenos, decaindo com o logaritmo do tamanho da rede, $L \approx \ln N / \ln \langle k \rangle$, sendo $\langle k \rangle = 2M/N = p(N - 1)$ o número médio de arestas na rede.

Diversos sistemas na natureza e na sociedade, tais como redes de reações químicas, redes de roteadores e a de Internet, foram inicialmente representados como grafos aleatórios. Contudo, estudos posteriores concluíram que a topologia e evolução de redes reais são governadas por princípios de organização robustos e que as mesmas não poderiam ser representadas por um modelo de rede puramente aleatório (ALBERT; BARABÁSI, 2002).

2.1.4.2 O modelo de Watts e Strogatz (1998)

Watts e Strogatz (1998) observaram que a conexão topológica em redes reais, incluindo redes biológicas, sociais e tecnológicas, não era completamente regular ou completamente aleatória, todavia se situava entre esses dois extremos. Ou seja, que redes reais poderiam ser altamente agrupadas, como redes regulares, mas com características de pequeno caminho, como grafos aleatórios. Esta característica é conhecida como a propriedade “mundo pequeno” (em inglês, *small world*). O conceito de redes pequeno mundo em termos simples descreve o fato de que apesar de seu tamanho amplo, na maior parte das redes há um menor caminho entre quaisquer dois vértices. A manifestação mais popular das redes pequeno mundo é a “separação de grau seis”, descoberta pelo psicólogo Milgran (1967), que concluiu que há em média um caminho de comprimento de seis pessoas entre a maioria de pares de pessoas nos Estados Unidos.

Baseados nesta descoberta, Watts e Strogatz sugeriram um modelo alternativo ao modelo de grafos aleatórios, conhecido como o *Modelo Small World de Watts-*

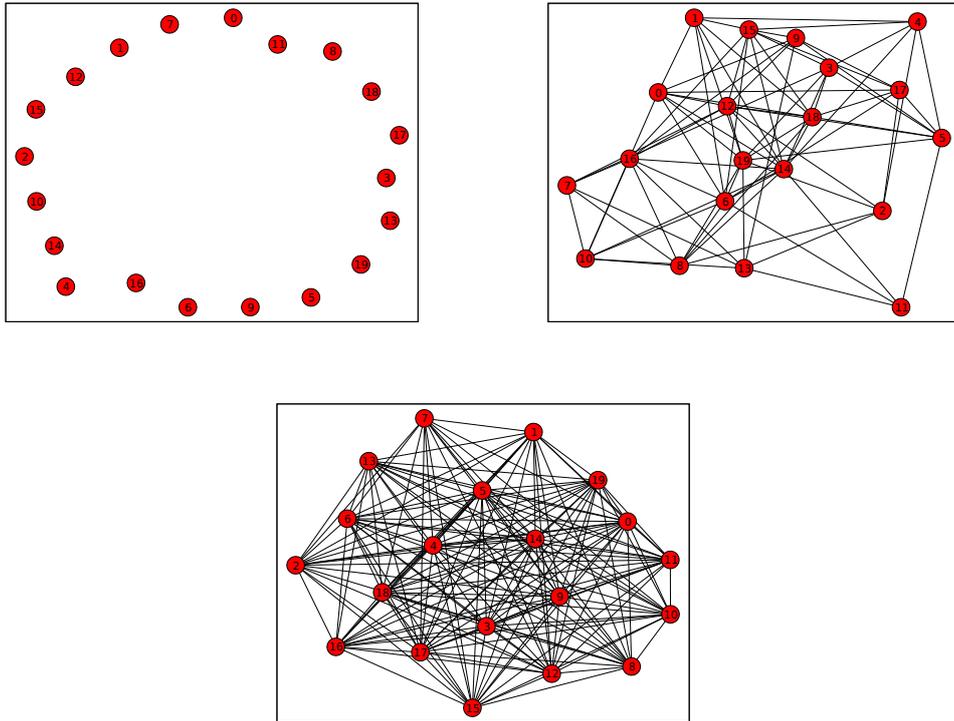


Figura 2.7 - Redes geradas a partir do modelo de Erdős e Rényi utilizando valores de p iguais a 0.0; 0.5 e 1.0, respectivamente. Com o incremento de p a rede torna-se cada vez mais conectada até que para $p = 1.0$ todos os vértices estão conectados uns aos outros.

Strogatz. A construção deste modelo é feita através de um processo de interpolação entre uma rede regular e uma rede aleatória, obtida a partir do seguinte processo:

- Inicia-se com uma rede circular e regular de N vértices, cada qual conectado aos seus k vizinhos mais próximos.
- Escolhe-se um vértice e uma aresta que o conecta a seu vizinho mais próximo em um sentido horário. Com probabilidade p , reconecta-se esta aresta a um vértice escolhido aleatoriamente (de maneira uniforme) ao longo da rede, com proibição de duplicações ou auto-conexões. Em outras palavras, muda-se a aresta de lugar.
- Repete-se este processo sobre todos os vértices da rede até que uma volta seja completada.

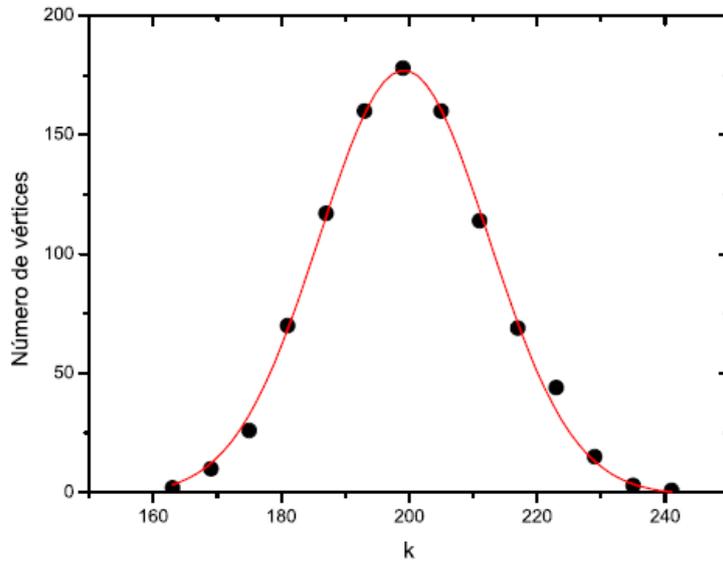


Figura 2.8 - Distribuição dos graus de conectividade de uma rede gerada a partir do modelo de Erdős e Rényi com 1,000 vértices, conectados com uma probabilidade $p = 0.2$.

Fonte: (RODRIGUES, 2007)

- Considera-se agora as arestas que conectam vértices a seus segundos vizinhos mais próximos. Como anteriormente, reescreve-se aleatoriamente cada uma dessas arestas com probabilidade p , e continua-se este processo, circulando sobre a rede e indo em direção aos vizinhos mais distantes após cada volta, até que cada ligação na rede original tenha sido considerada uma vez. Como há $Nk/2$ arestas em toda a rede, o processo de reescrita pára após $k/2$ voltas.

A Figura 2.9 mostra as redes obtidas a partir do processo descrito acima, para três diferentes realizações. O número de vértices em todos os casos é $N = 20$ e o número de vizinhos iniciais é $k = 4$. Pode-se observar que para $p = 0.0$, a rede original é regular e a mesma se mantém inalterada; com o incremento de p a rede torna-se cada vez mais desordenada até que para $p = 1.0$ todas as arestas são reescritas aleatoriamente.

A Figura 2.10 mostra os valores do coeficiente de agrupamento C e do comprimento do caminho médio L , ambos em função do parâmetro p , com $0.0 \leq p \leq 1.0$. Tais valores foram normalizados pelos valores de $C(0)$ e $L(0)$, calculados a partir de redes

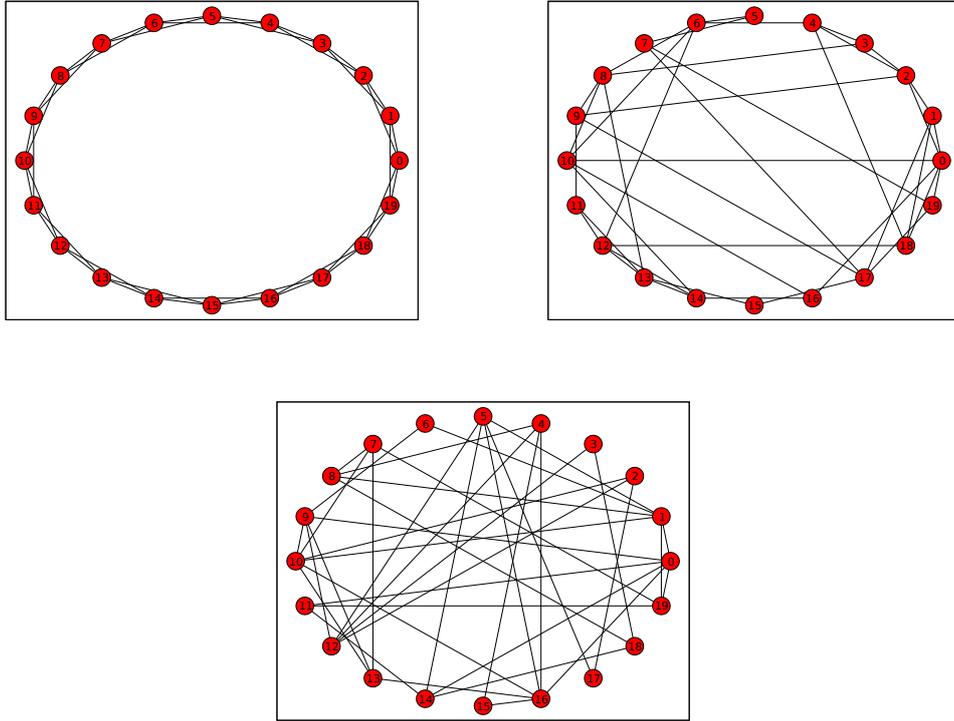


Figura 2.9 - Procedimento de interpolação entre uma rede regular e uma rede aleatória, sem alterar o número de vértices ou arestas. Uma rede regular é obtida quando $p = 0.0$ e o valor $p = 1.0$ produz uma rede aleatória. Valores intermediários de p como, por exemplo, $p = 0.5$ produzem redes do tipo pequeno mundo.

regulares ($p = 0.0$), respectivamente. O número de vértices é $N = 1,000$, o número de vizinhos iniciais é $k = 10$ e o eixo- x está em escala logarítmica.

Pode-se observar que a rede regular ($p = 0.0$) é altamente agrupada e possui o comprimento do caminho médio elevado. Contudo, a rede aleatória ($p = 1.0$) é fracamente agrupada e possui o comprimento do caminho médio reduzido. Mais importante, a Figura 2.10 revela um intervalo de valores de p sobre o qual $L(p)$ é praticamente pequeno como L_{random} e ainda $C(p) \gg C_{random}$, onde L_{random} e C_{random} são o comprimento do menor caminho e o coeficiente de agrupamento associados a uma rede aleatória, respectivamente. Estas redes do tipo pequeno mundo resultam de uma queda abrupta de $L(p)$ devido a introdução de poucas arestas de longo alcance, o que diminui consideravelmente a distância entre os vértices. Em resumo, no fenômeno pequeno mundo tem-se: $L \gtrsim L_{random}$ mas $C \gg C_{random}$.

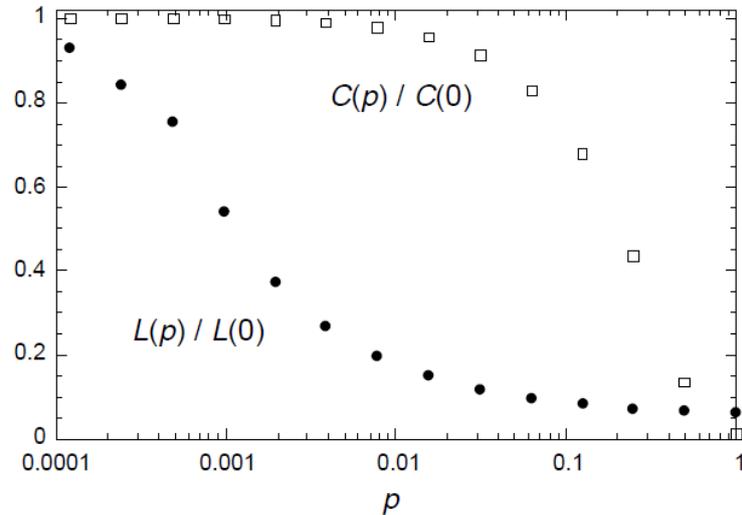


Figura 2.10 - Valores de C e L em função do parâmetro p .
 Fonte: (WATTS; STROGATZ, 1998)

A Figura 2.11 mostra a distribuição do grau de conectividade para uma rede pequeno mundo formada por 1,000 vértices. Pode-se observar que este histograma segue aproximadamente uma distribuição de Poisson, o que significa que, em média, a maior parte dos vértices tem o mesmo grau de conectividade e que poucos vértices possuem grau de conectividade inferior ou superior à média.

2.1.4.3 O modelo de Barabási e Albert (1999)

Barabási e Albert (1999) mapearam a topologia das conexões entre as páginas da Teia Mundial e descobriram que além da mesma apresentar o fenômeno pequeno mundo, sua distribuição de conexões não era aleatória, mas do tipo *livre de escala*, ou seja, da forma $P(k) \approx k^{-\gamma}$.

A distribuição livre de escala é um tipo de distribuição de probabilidade que reflete a invariância de escala. Este tipo de distribuição possui uma lei de potência associada, que é uma curva continuamente decrescente sem um pico característico, sendo descrita por um único expoente. As quantidades geradas aleatoriamente possuem uma escala típica, sendo descritas por curvas características definidas por uma média e um desvio padrão. Diferentemente da uniformidade, as leis de potência sugerem que muitos eventos pequenos podem coexistir com poucos eventos grandes.

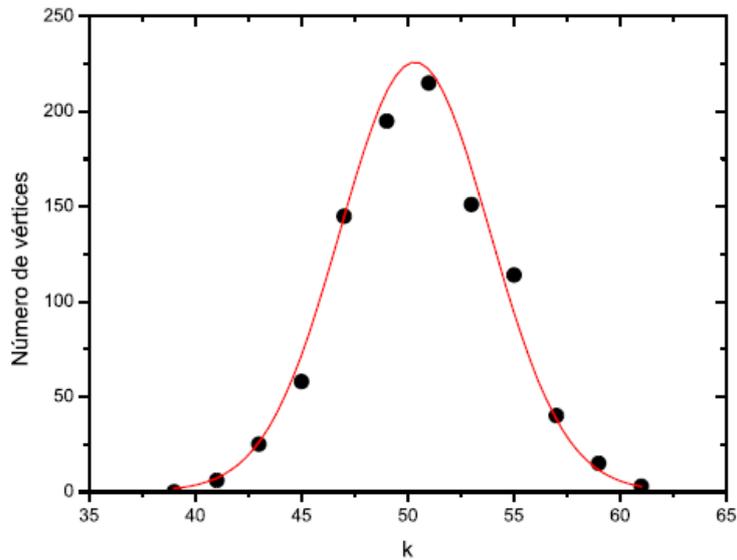


Figura 2.11 - Distribuição do grau de conectividade para uma rede pequeno mundo formada por 1,000 vértices, $k = 25$ vizinhos próximos e $p = 0.3$.
 Fonte: (RODRIGUES, 2007)

Barabási e Albert propuseram um modelo de crescimento capaz de gerar redes livres de escala, baseado nos seguintes passos:

- *Crescimento*: Inicia-se uma rede com um pequeno número de vértices N_0 . A cada passo é adicionado um novo vértice com M arestas ($M \leq N_0$) que se conectam com os vértices já presentes na rede.
- *Ligação preferencial*: O novo vértice a ser adicionado na rede tende a se conectar com os vértices com maior grau de conectividade, ou seja, a probabilidade de um vértice n_j ser escolhido é proporcional ao seu grau de conectividade.

Pode-se observar que os dois mecanismos descritos acima para a construção das redes livres de escala não estão presentes no modelo aleatório de Erdős e Rényi e no modelo de redes pequeno mundo de Watts e Strogatz, já que nos mesmos não há adição de novos vértices e as arestas são estabelecidas de forma homogênea. Por outro lado, nas redes geradas pelo modelo livre de escala, os vértices mais conectados tendem a receber mais arestas. As redes resultantes são então formadas por um número reduzido de vértices altamente conectados, denominados *hubs*, e por uma

grande quantidade de vértices poucos conectados, o que define a distribuição livre de escala.

A Figura 2.12 mostra as redes obtidas a partir do processo descrito acima, para redes com 20, 30 e 50 vértices e $M = 1$. Pode-se observar que, em todos os casos, as redes geradas apresentam um número reduzido de hubs. A Figura 2.13 mostra a distribuição do grau de conectividade para uma rede livre de escala com 10,000 vértices.

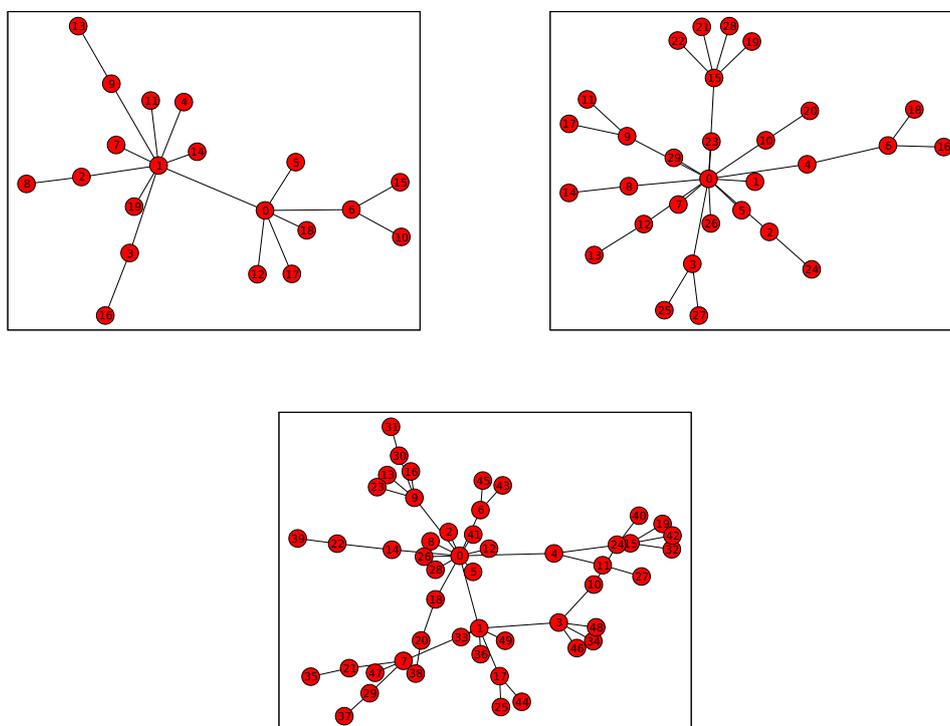


Figura 2.12 - Redes geradas a partir do modelo de Barabási e Albert com 20,30 e 50 vértices e $M = 1$, respectivamente. Em todos os casos pode-se observar a presença de um número reduzido de hubs.

As representações gráficas das redes presentes neste capítulo foram obtidas por meio do software NetworkX, que é um módulo pertencente à linguagem Python para a criação, manipulação e estudo da estrutura e dinâmica de redes complexas. O NetworkX é software de acesso livre, em código aberto, e compatível com o sistema operacional Windows, Linux e Mac.

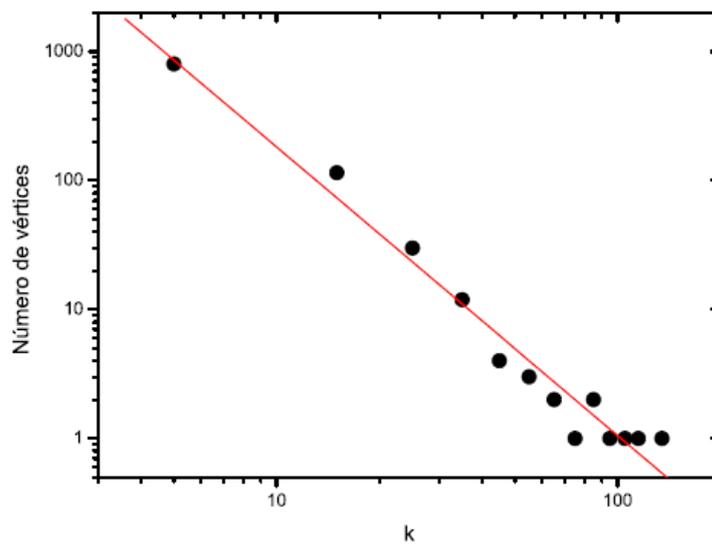


Figura 2.13 - Distribuição do grau de conectividade para uma rede livre de escala formada por 10,000 vértices e $M = 5$. Esta distribuição segue uma lei de potência da forma $P(k) \approx k^{-\gamma}$.

Fonte: (RODRIGUES, 2007)

2.2 Análise de séries temporais

2.2.1 Introdução

Um *sistema dinâmico* é um sistema que evolui a cada instante de acordo com um conjunto de regras fixas que determinam como um estado do sistema se altera para um outro. Dois tipos principais de sistemas dinâmicos são encontrados em aplicações: aqueles em que a variável tempo é contínua ($t \in \mathbb{R}$) e aqueles em que a variável tempo é discreta ($t \in \mathbb{N}$).

Um *sistema dinâmico contínuo* pode ser descrito por equações diferenciais ordinárias (EDO's) ou parciais (EDP's). Para o caso de uma ou mais EDO's, tal sistema assume a forma

$$\begin{aligned} d\vec{x}(t)/dt &= \vec{f}_\mu(\vec{x}(t)) \\ \vec{x}(0) &= \vec{x}_0 \end{aligned} \quad (2.12)$$

enquanto um *sistema dinâmico discreto* pode ser representado como a iteração de uma ou mais funções, isto é

$$\vec{x}_{n+1} = \vec{f}_\mu(\vec{x}_n), \quad (2.13)$$

onde $\vec{x} \in \mathbb{R}^m$ corresponde ao vetor de estados do sistema, \vec{f}_μ constitui um conjunto de m funções, μ é definido como um *parâmetro de controle* do sistema e $\vec{x}(0)$ é o vetor de condições iniciais das variáveis de estado do sistema dinâmico.

O conjunto dos valores assumidos pelas variáveis de estado ao longo do tempo, a partir da condição inicial $\vec{x}(0)$, é chamada de *trajetória* ou *órbita* do sistema dinâmico. As trajetórias descritas pelas variáveis de estado de um sistema dinâmico são usualmente representadas em um espaço euclidiano \mathbb{R}^m , que neste contexto é chamado de *espaço de estados* ou ainda *espaço de fase*, sendo m a dimensão do sistema. Se \vec{f}_μ é constituído de m funções contínuas lineares, o sistema dinâmico é linear; caso contrário, o sistema dinâmico é não-linear. Se \vec{f}_μ não depende explicitamente do tempo t , o sistema dinâmico é autônomo; caso contrário o sistema é não-autônomo (ALLIGOOD et al., 1997).

Um exemplo de sistema dinâmico contínuo, não-linear e autônomo de EDO's é dado por:

$$\begin{aligned} dx/dt &= \sigma(-x + y) \\ dy/dt &= rx - y - xz \\ dz/dt &= -bz + xy \end{aligned} \quad (2.14)$$

onde σ , r e b são parâmetros reais e positivos.

O sistema de equações (2.14) é comumente denominado sistema de Lorenz (LORENZ, 1963) e fornece uma aproximação para a dinâmica de uma camada de fluido convectiva. A solução deste sistema pode apresentar comportamento estável, periódico ou quasiperiódico. Para valores de parâmetros iguais a $\sigma = 10$, $b = 8/3$ e $r = 28$, por exemplo, a solução deste sistema oscila confinada em uma região limitada do espaço de fase, sem entretanto, apresentar periodicidade. Tal trajetória é *atraída* para um sub-conjunto de pontos do espaço de fase, conhecido como *atrator caótico* (ALLGOOD et al., 1997).

Quando as equações que governam um sistema dinâmico são conhecidas, o estudo das características de suas soluções pode revelar comportamentos complexos e interessantes. Porém, em situações reais raramente se dispõe de um conjunto de equações diferenciais ou mesmo um mapa que descreva o comportamento do sistema. Em geral, monitora-se uma única variável a partir de um experimento que se sabe de antemão depender de outras variáveis (KANTZ; SCHREIBER, 2003). Desta forma, obtém-se uma série temporal de medidas:

$$X = \{x(t) | t \in \mathbb{N}, x(t) \in \mathbb{R}\}, \quad (2.15)$$

com T pontos. São exemplos de séries temporais a evolução das variáveis de estado $x(t)$, $y(t)$ e $z(t)$, em função do tempo t do Sistema (2.14), conforme mostra a Figura 2.14.

2.2.2 Técnicas de Análise de Séries Temporais

2.2.2.1 Análise espectral

As técnicas de análise espectral são utilizadas, em geral, na detecção de periodicidades ou padrões cíclicos em séries temporais. Sob condições bem gerais, uma função $f(t)$ pode ser considerada como sendo a superposição de um número (eventualmente infinito) de componentes periódicas. A determinação do peso relativo de cada uma dessas componentes é chamada de *análise espectral*. Se $f(t)$ é periódica, seu espectro pode ser representado como a combinação linear de oscilações cujas frequências são múltiplos inteiros da frequência básica ω . Essa combinação linear é chamada *série de Fourier*. Quando $f(t)$ é não-periódica, o espectro de frequências varia continuamente

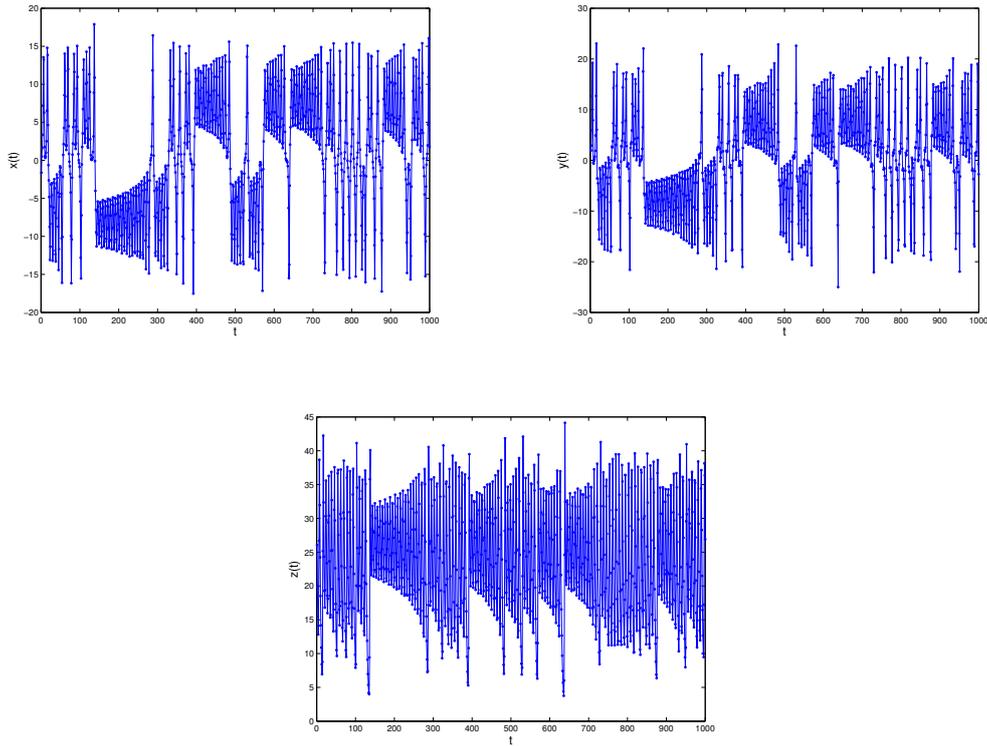


Figura 2.14 - Exemplos de séries temporais obtidas a partir da evolução das variáveis de estado $x(t)$, $y(t)$ e $z(t)$ em função do tempo t do Sistema (2.14), respectivamente.

e usa-se a chamada *transformada de Fourier* para representar $f(t)$ em termos dessas frequências (PAPOULIS, 1962). Escreve-se a transformada de Fourier de $f(t)$ como:

$$f(\omega) = \int_{-\infty}^{+\infty} e^{-i\omega t} f(t) dt. \quad (2.16)$$

O espectro de potências $P(\omega)$, que indica o “peso relativo” com que a frequência ω comparece na composição de $f(t)$, é definido como o quadrado do módulo de $f(\omega)$, ou seja,

$$P(\omega) = |f(\omega)|^2. \quad (2.17)$$

Na situação do presente trabalho, dispõe-se de uma série temporal finita e discreta da forma (2.15). Se T é o número total de pontos na série, então x_t corresponde a um tempo total de medida $t_{\max} = T\Delta t$. A transformada discreta de Fourier, de uma

série temporal, é definida por uma outra série \hat{x}_k tal que:

$$\hat{x}_k = \frac{1}{\sqrt{T}} \sum_{t=1}^T x_n \exp \left[i \frac{2\pi tk}{T} \right], \quad k = 1, \dots, T. \quad (2.18)$$

A série temporal (2.15) depende do tempo; por outro lado, \hat{x}_k depende das frequências, isto é, $\hat{x}_k = \hat{x}(\omega = k\Delta f)$ com $\Delta f = 1/t_{\max}$. O cálculo da transformada discreta de Fourier (2.18) pode ser feito com rapidez através do uso do algoritmo conhecido como FFT (*Fast Fourier Transform*) (COOLEY; TUKEY, 1965).

O espectro de potências $P(\omega)$ para uma série temporal discreta é definido por:

$$P(\omega) = |\hat{x}_k|^2. \quad (2.19)$$

A Figura 2.15 apresenta os espectros de potências de um ruído branco uniformemente distribuído entre $[0, 1]$, de uma série periódica (função seno) e da série temporal $x(t)$ que é parte da solução do Sistema (2.14). Séries temporais com evoluções diferentes apresentam espectros de potência distintos. Sinais aleatórios apresentam espectros de potência com energia constante ao longo das frequências. Sinais periódicos de período P apresentam um pico bem definido na frequência correspondente a esse período. Por outro lado, séries temporais quasiperiódicas apresentam espectros de potência com uma “banda larga”, o que indica a existência de um contínuo de frequências.

2.2.2.2 Função de autocorrelação

Uma importante medida para se identificar as propriedades de uma série temporal consiste no cálculo da função de autocorrelação $A_{xx}(\tau)$, definida como:

$$A_{xx}(\tau) = \frac{1}{T - \tau} \sum_{t=1}^{T-\tau} \frac{(x_t - \bar{x})(x_{t+\tau} - \bar{x})}{s_{xx}^2}, \quad (2.20)$$

onde T é o número de pontos da série X , \bar{x} sua média e s_{xx}^2 sua variância. Essa função representa a média do produto dos valores da série temporal nos instantes t e $t + \tau\Delta t$ e indica por quanto tempo o valor da série temporal no instante t depende de seus valores prévios; em outras palavras $A_{xx}(\tau)$ mede o grau de semelhança existente no sinal à medida que o tempo passa (BAKER; GOLLUB, 1996).

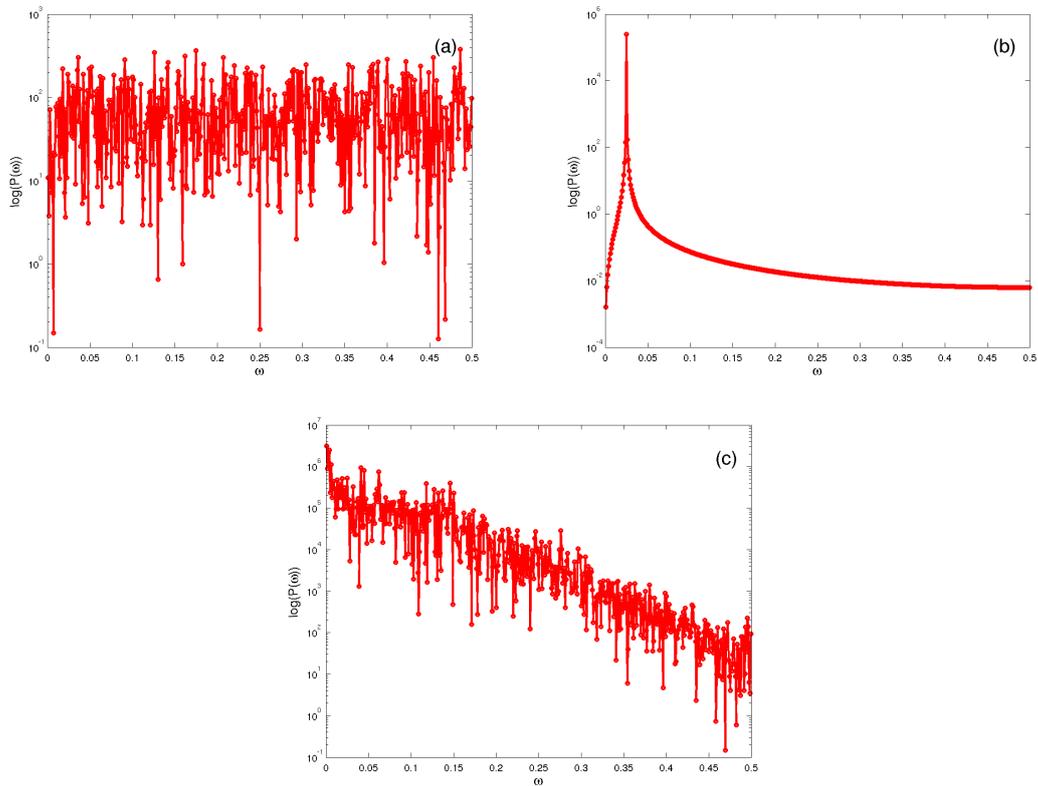


Figura 2.15 - Espectros de potências de séries temporais com dinâmicas distintas. (a) Série temporal aleatória uniformemente distribuída entre $[0, 1]$. (b) Série periódica (função seno). (c) Série temporal quasiperiódica $x(t)$, oriunda do sistema de Lorenz (2.14).

A função de autocorrelação (2.20) pode ser utilizada na descrição das propriedades de uma série temporal. Se a série temporal é um ruído branco, por exemplo, $A_{xx}(\tau)$ tem valor unitário quando $\tau = 0$ e é aproximadamente nula para todos os demais valores de τ . $A_{xx}(\tau)$ de um sinal periódico é igualmente periódica, pois o sinal periódico volta a se parecer consigo mesmo após um intervalo de tempo correspondente ao período. Por fim, em uma série quasiperiódica $A_{xx}(\tau) \rightarrow 0$ quando $\tau \rightarrow \infty$; a semelhança de uma série temporal consigo mesma diminui com o tempo e acaba por desaparecer completamente (Fig. 2.16) (ARGYRIS et al., 1994).

Analogamente, o grau de semelhança entre duas séries temporais $X = \{x(t)|t \in \mathbb{N}, x(t) \in \mathbb{R}\}$ e $Y = \{y(t)|t \in \mathbb{N}, y(t) \in \mathbb{R}\}$, com T pontos cada, pode ser obtido através da função de correlação cruzada $C_{xy}(\tau)$, definida como:

$$C_{xy}(\tau) = \frac{1}{T - \tau} \sum_{t=1}^{T-\tau} \frac{(x_t - \bar{x})(y_{t+\tau} - \bar{y})}{s_x s_y}, \quad (2.21)$$

onde \bar{x} e \bar{y} são as médias associadas as séries X e Y e s_x e s_y os desvios padrões de X e Y , respectivamente.

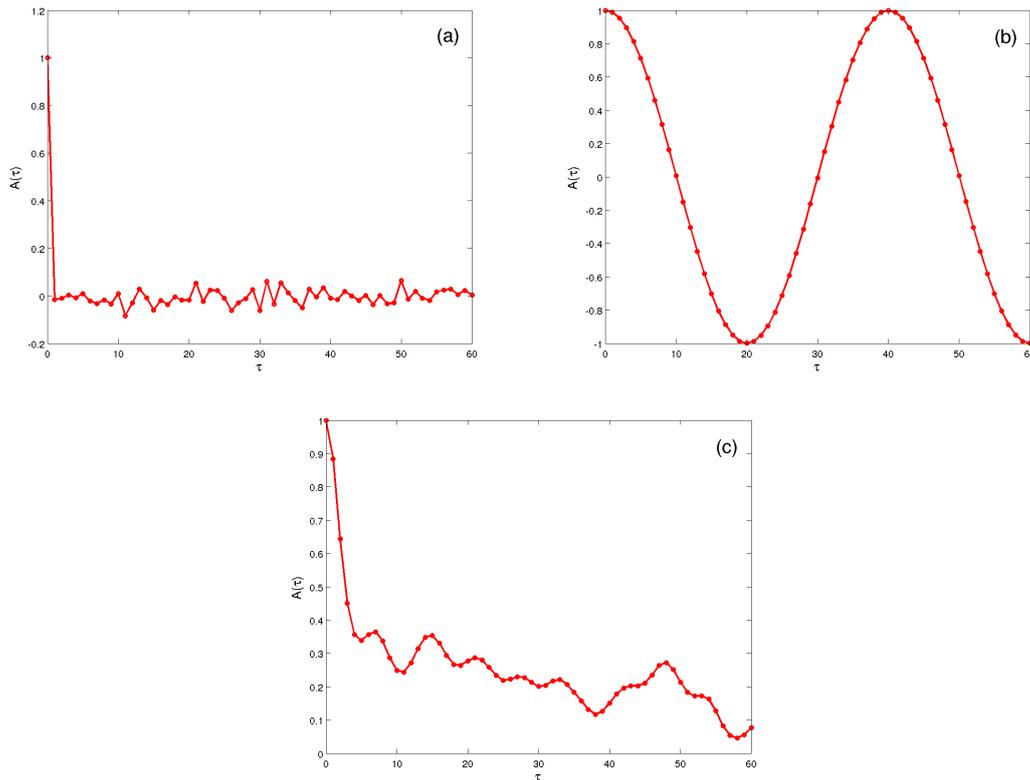


Figura 2.16 - Funções de autocorrelação de séries temporais com dinâmicas distintas com $0 \leq \tau \leq 60$. (a) Ruído branco uniformemente distribuído entre $[0, 1]$. (b) Série periódica (função seno). (c) Série temporal $x(t)$ do sistema de Lorenz (2.14).

2.2.2.3 Coeficiente de Hurst

Hurst (HURST, 1951) propôs uma técnica estatística conhecida como Análise R/S (em inglês, R/S analysis) para a detecção de correlação (persistência) em uma série temporal. Esta medida é definida como:

$$\frac{R_T}{S_T} = \frac{1}{s_T} \left[\max_{1 \leq k \leq T} \sum_{j=1}^k (x_j - \bar{x}) - \min_{1 \leq k \leq T} \sum_{j=1}^k (x_j - \bar{x}) \right], \quad (2.22)$$

onde T é o número de pontos de X e s_T seu desvio padrão. Em geral, observa-se como R/S varia com sucessivos intervalos τ de T substituindo-se T por τ na Equação (2.22), e então:

$$\left(\frac{R_\tau}{S_\tau} \right)_{\text{av}} = \left(\frac{\tau}{2} \right)^{H_u}. \quad (2.23)$$

O valor do coeficiente de Hurst é então obtido através de um ajuste linear da curva de $\log(R_\tau/S_\tau)_{\text{av}}$ contra $\log(\tau/2)$. Os valores de H_u variam entre 0.0 e 1.0. O valor de $H_u = 0.5$ indica ausência de correlação (memória). Anti-persistência (anti-correlação) é caracterizada por um valor de H_u pertencente ao intervalo $0.0 \leq H_u < 0.5$ e persistência (correlação) por um valor de H_u pertencente ao intervalo $0.5 < H_u \leq 1.0$.

2.2.3 Aproximação por Redes Complexas

De maneira geral, as técnicas em análise de séries temporais permitem pesquisadores resumir as características de uma série temporal, e desta forma, entender a dinâmica subjacente de um sistema ou prever como o mesmo evolui com o tempo. Novas técnicas têm sido propostas para capturar informações adicionais ou quantificar séries temporais de novas formas (ZHANG et al., 2006; LAI et al., 2010; VERPLANCKE et al., 2010; AO, 2010). Um dos avanços mais interessantes foi sugerido pelo trabalho de Zhang & Small (ZHANG; SMALL, 2006) que utilizou, pela primeira vez, a teoria de redes complexas na análise de séries temporais.

2.2.3.1 O trabalho de Zhang e Small (2006)

A idéia proposta no referido trabalho consiste em dividir uma série temporal em n ciclos disjuntos de acordo com seus mínimos locais (ou máximo locais), denotados

por $\{C_1, C_2, \dots, C_n\}$, sendo que cada ciclo corresponde um vértice da rede associada. A conectividade entre os vértices é obtida com base no coeficiente de correlação linear ρ entre cada par de ciclos, dado por:

$$\rho = \frac{C_{xy}}{\sqrt{A_{xx}}\sqrt{A_{yy}}} \quad , \quad (2.24)$$

onde A_{xx} e C_{xy} correspondem à autocorrelação (2.20) e à correlação cruzada (2.21) com $\tau = 0$, respectivamente. O coeficiente de correlação caracteriza a similaridade entre os ciclos, assim pares de ciclos com uma correlação temporal superior a um limite pré-estabelecido D , estarão conectados na rede correspondente.

Tais estudiosos trabalharam com dois tipos de séries pseudo-periódicas: o primeiro tipo se refere a uma série periódica acrescida com uma taxa de ruído (denominada “série periódica com ruído”), ou seja:

$$y_n = \sin(2\pi\omega n) + b\eta_n, \quad (2.25)$$

onde $b = 0.2836$ e η é um parâmetro aleatório tal que $\eta \approx N(0, \sigma^2)$.

O segundo tipo se refere a série temporal oriunda da evolução temporal da variável de estado x do sistema de equações diferenciais ordinárias de Rossler (denominada “série caótica”):

$$\begin{aligned} dx/dt &= -y - z \\ dy/dt &= x + ay \\ dz/dt &= d + (x - c)z \end{aligned} \quad (2.26)$$

onde a , d e c são parâmetros reais e positivos.

A representação utilizada no referido trabalho codifica a dinâmica da série temporal em estudo em uma topologia de rede complexa, e desta forma, séries temporais com dinâmicas diferentes exibem estruturas topológicas de rede distintas (conforme Figura 2.17). Especificamente, séries temporais periódicas com ruído correspondem a redes aleatórias, e séries temporais caóticas geram redes com características de redes pequeno mundo e livre de escala. Estas classificações foram obtidas através do estudo das propriedades topológicas usuais de uma rede, como a distribuição do grau de conectividade $P(k)$ versus k , o comprimento do caminho médio e o coeficiente de

agrupamento.

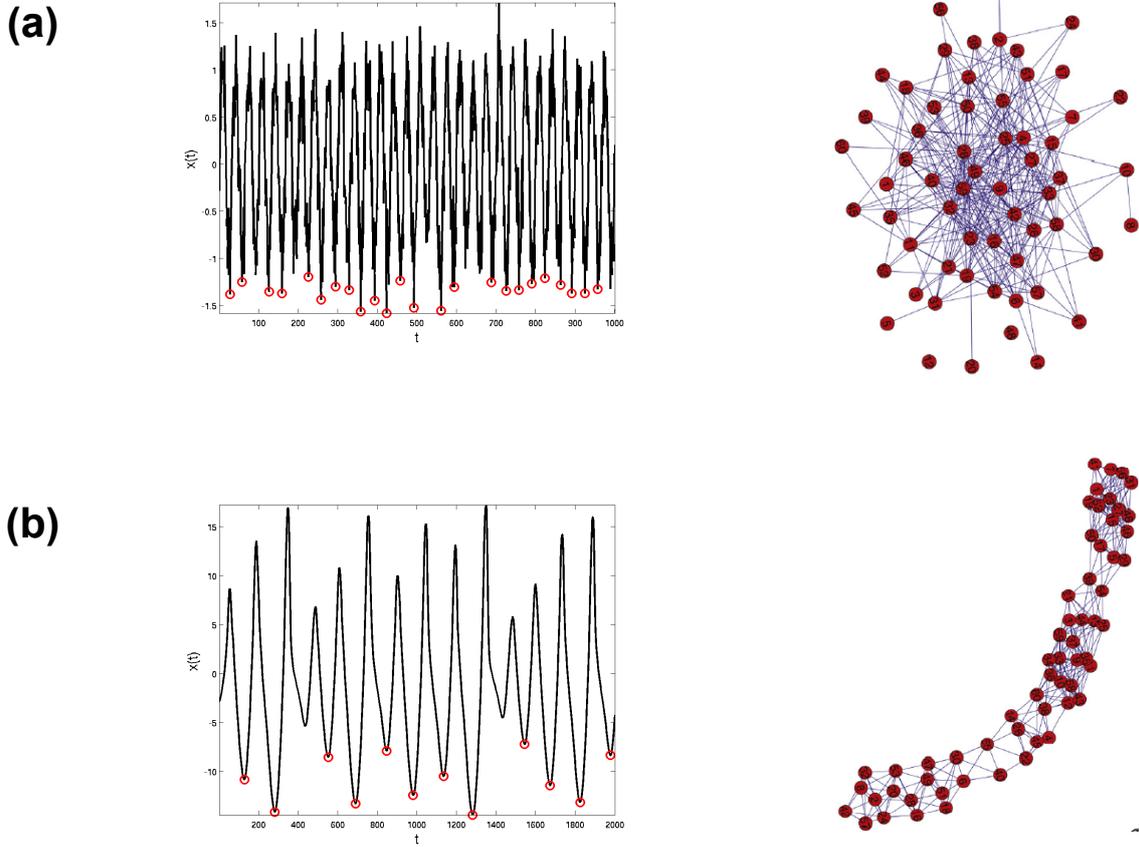


Figura 2.17 - Séries temporais pseudo-periódicas (com mínimos locais representados por círculos, em vermelho) e suas representações em redes complexas. (a) A série temporal aleatória acrescida a uma taxa de ruído está associada a uma rede aleatória. (b) A série temporal caótica está associada a uma rede com características de redes pequeno mundo e livre de escala.

Fonte: Figura adaptada de Zhang e Small (2006).

O mapeamento proposto por Zhang e Small (2006) identifica com êxito a dinâmica de uma série temporal a partir do estudo de sua rede correspondente. Porém, este procedimento faz uso de ciclos presentes em uma série temporal, e desta forma, o mesmo fica restrito à análise de séries temporais com periodicidade. Além disso, as topologias das redes geradas através deste mapeamento são altamente sensíveis aos valores assumidos pelo parâmetro D . Ou seja, valores de D próximos de 1.0 geram redes altamente esparsas, enquanto que valores de D próximos de 0.0 geram redes

altamente densas.

2.2.3.2 O trabalho de Lacasa et al. (2008)

Posteriormente, Lacasa et al. (2008) propuseram uma estratégia de mapeamento de uma série temporal em uma rede complexa baseada no conceito de *visibilidade*. Para ilustrar a idéia proposta, a Figura 2.18 apresenta os vinte primeiros valores de uma série temporal periódica (parte superior) como também a representação desta série em um gráfico de barras (parte intermediária). Uma dada barra é conectada com todas aquelas que podem ser vistas do topo da primeira, obtendo-se assim o grafo associado (parte inferior). Neste grafo, cada vértice corresponde, na mesma ordem, a um ponto da série temporal, e dois vértices serão conectados se existe visibilidade entre os pontos correspondentes da série temporal, ou seja, se existe uma linha que os conecta, de tal forma que a “linha de visibilidade” não intercepta nenhum ponto intermediário.

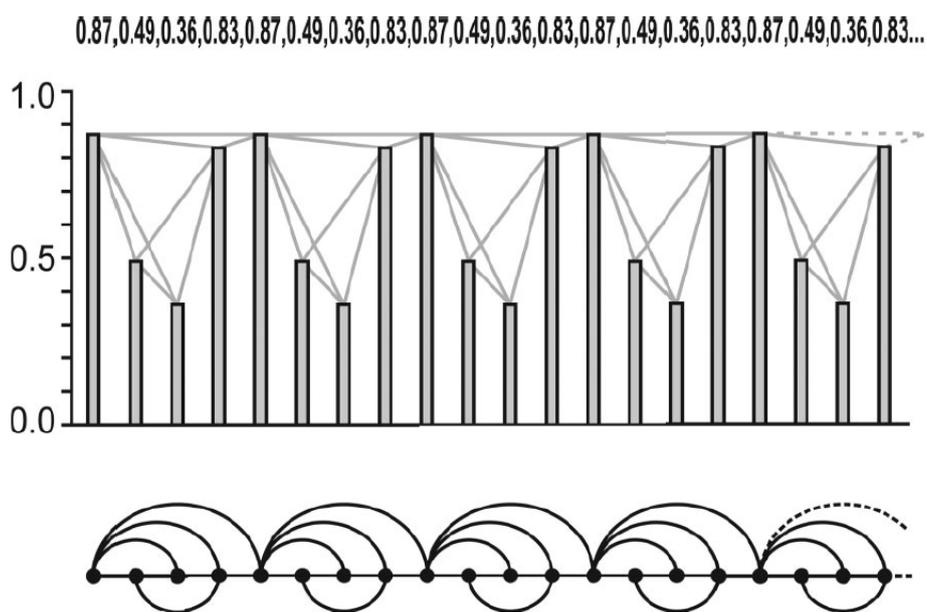


Figura 2.18 - Exemplo de uma série temporal com 20 pontos e sua respectiva rede complexa, obtida a partir do algoritmo de visibilidade descrito anteriormente. Nesta rede cada vértice corresponde, na mesma ordem, a um ponto da série temporal. A visibilidade entre os pontos define a ligação entre os vértices na rede associada.

Fonte: (LACASA et al., 2008)

Formalmente, é possível estabelecer o seguinte critério de visibilidade entre os pontos de uma série temporal: dois pontos arbitrários de uma série temporal (t_a, y_a) e (t_b, y_b) serão visíveis, e conseqüentemente, serão vértices conectados na rede associada, se qualquer outro ponto da série temporal (t_c, y_c) colocado entre eles satisfizer a relação:

$$y_c \leq y_b + (y_a - y_b) \frac{t_b - t_c}{t_b - t_a} . \quad (2.27)$$

O processo de mapeamento desenvolvido por [Lacasa et al. \(2008\)](#) foi aplicado em três conjuntos distintos de séries temporais. O primeiro conjunto se refere a séries periódicas com diferentes períodos. Neste caso, todas as séries temporais foram mapeadas em redes regulares (como aquela obtida na parte inferior da Figura 2.18). O segundo conjunto se refere a séries aleatórias obtidas a partir de uma distribuição uniforme entre $[0, 1]$. Neste caso, todas as séries temporais aleatórias foram mapeadas em redes aleatórias. Por fim, o último conjunto analisado se refere a séries fractais, que foram mapeadas em redes do tipo pequeno mundo.

Analogamente ao mapeamento proposto por [Zhang e Small \(2006\)](#), o mapeamento proposto por [Lacasa et al. \(2008\)](#) codifica a dinâmica da série temporal em estudo em uma topologia de rede complexa. Este mapeamento é mais amplo se comparado com o primeiro, já que sua aplicação não está restrita à análise de séries temporais com periodicidade.

Após os trabalhos de [Zhang e Small \(2006\)](#) e [Lacasa et al. \(2008\)](#), diversos mapeamentos de séries temporais em redes complexas foram sugeridos ([LUQUE et al., 2009](#); [YANG; YANG, 2008](#); [MARWAN et al., 2009](#); [XU. et al., 2008](#); [GAO; JIN, 2009](#)) com base em conceitos como visibilidade ([LUQUE et al., 2009](#)), correlações ([YANG; YANG, 2008](#)), análise de recorrência ([MARWAN et al., 2009](#)) e reconstrução do espaço de fase ([XU. et al., 2008](#); [GAO; JIN, 2009](#)). Estes estudos demonstraram que, independentemente do mapeamento utilizado, séries temporais com diferentes características são mapeadas em redes com propriedades topológicas distintas. Este fato sugere que pode ser possível diferenciar as propriedades de uma série temporal utilizando medidas em redes complexas, contudo, não é ainda claro como a análise de séries temporais pode ser útil na investigação das propriedades topológicas de uma rede complexa.

3 MAPEAMENTO PROPOSTO

Seja \mathcal{M} um mapeamento de uma série temporal contínua $X \in \mathcal{T}$ em uma rede complexa $g \in \mathcal{G}$. Em condições ideais, tal mapeamento preservaria *toda* a informação da série original, possivelmente através de um mapa bijetivo \mathcal{M}_b onde cada série temporal X seria mapeada exatamente em uma rede complexa $g = \mathcal{M}_b[X]$ que, por sua vez, seria inversivelmente mapeada na mesma série temporal $X = \mathcal{M}_b^{-1}[\mathcal{M}_b[X]]$. Na prática, isto se torna impossível já que séries temporais contínuas possuem infinitos valores enquanto redes complexas são limitadas a um conjunto finito de vértices e arestas. Desta forma, qualquer mapeamento de uma série temporal contínua X em uma rede complexa g deve discretizar de alguma forma a série temporal.

Neste trabalho, utiliza-se a discretização de X com base nos quantis de seus valores. Em estatística descritiva, o conceito de quantil é utilizado para particionar uma amostra com T pontos em Q subconjuntos (quantis) de mesmo tamanho, cada qual com T/Q pontos. Desta forma, o q_i -ésimo quantil de uma amostra, ordenada em ordem crescente, é obtido através da relação:

$$q_i = (T * i) / Q, \quad (3.1)$$

onde $i = 1, 2, \dots, Q$. Desta forma, dada uma série temporal X , primeiramente seus Q quantis são identificados, e então, cada quantil q_i é associado a um vértice $n_i \in \mathcal{N}$ na rede correspondente. Dois vértices n_i e n_j estarão conectados na rede com um arco $(n_i, n_j, w_{ij}) \in \mathcal{L}$, onde o peso w_{ij} de cada arco é dado pelo número de vezes que um dado ponto x_t no quantil q_i é seguido por um ponto x_{t+1} no quantil q_j (Fig. 3.1).

O mapeamento proposto, \mathcal{M}_{QT} , possui duas propriedades importantes. Primeiramente, ele é sobrejetivo. Dada uma série temporal X com T pontos e Q quantis, o mapeamento produzirá uma e somente uma rede complexa $g = \mathcal{M}_{QT}[X]$, contudo séries temporais distintas X e X' *podem* ser mapeadas em uma mesma rede complexa $g = \mathcal{M}_{QT}[X] = \mathcal{M}_{QT}[X']$. Segundo, o mapeamento direto é bastante simples já que o mesmo requer a especificação de um único parâmetro Q que pode ser obtido, por exemplo, através da relação $Q \approx \sqrt{T}$. Em diversos mapeamentos propostos (XU. et al., 2008; GAO; JIN, 2009; MARWAN et al., 2009; YANG; YANG, 2008), a estrutura da rede complexa resultante é bastante sensível à escolha de diversos parâmetros, tais como, tempo de atraso, dimensão de imersão e limiarização, necessários para a reconstrução do espaço de fases e análise de recorrência.

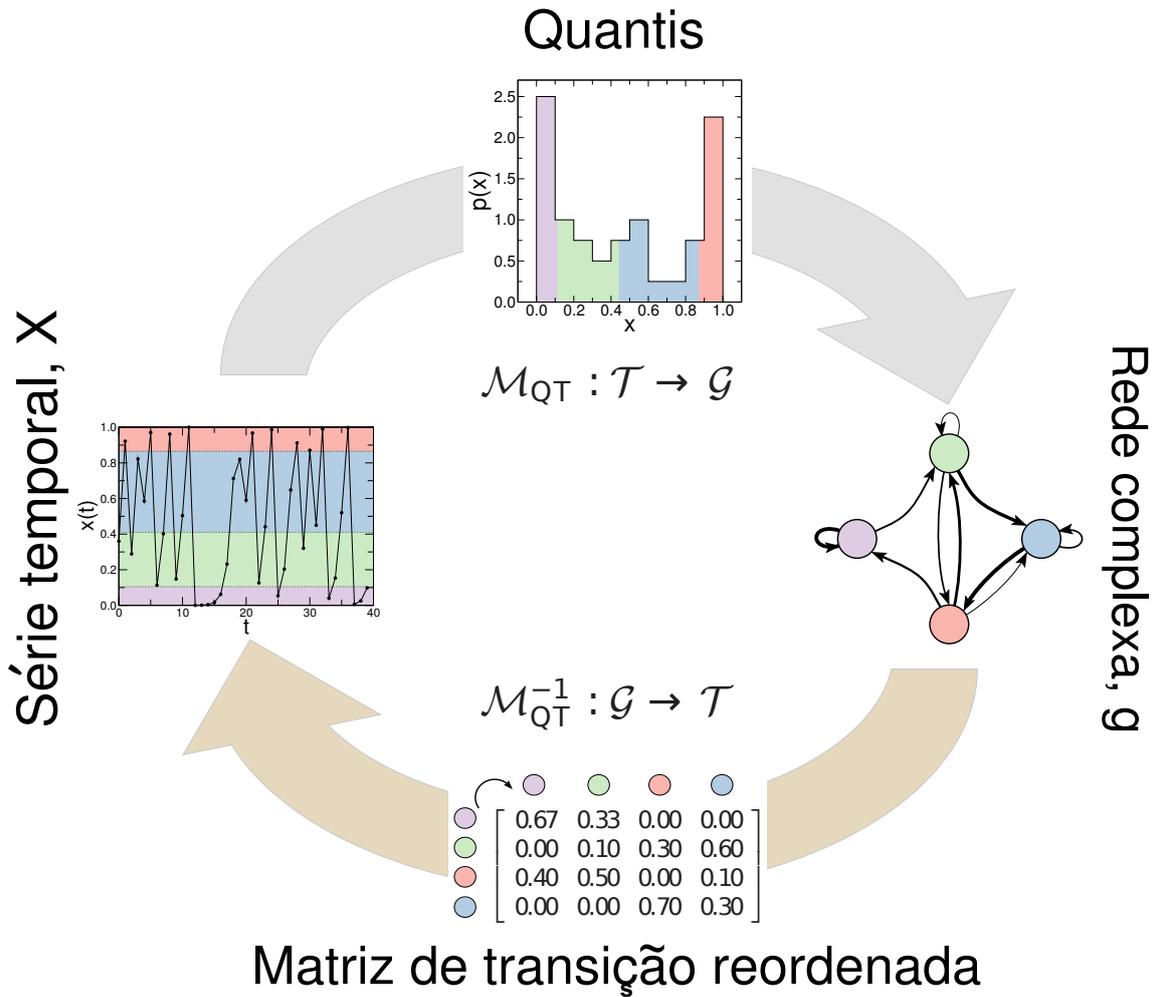


Figura 3.1 - Ilustração do mapeamento proposto. *Mapeamento direto*: Uma série temporal X é mapeada em uma rede complexa g com $Q = 4$ vértices (quantis). Transições repetidas entre quantis resultam em arestas na rede com pesos maiores (representados por linhas mais grossas). *Mapeamento inverso*: A partir de uma rede complexa g , constituída de $Q = 4$ vértices, é construída uma série temporal X movendo-se aleatoriamente T passos de um vértice n_i para um vértice n_j , com probabilidade w_{ij} .

O mapeamento proposto fornece uma vantagem significativa em relação aos mapeamentos existentes. O mesmo possui uma operação inversa aproximada \mathcal{M}_{QT}^{-1} . Esta operação consiste na realização de um caminho aleatório sobre a rede com probabilidade w_{ij} dada pela matriz de adjacência com pesos W , de tal forma que $\sum_j w_{ij} = 1$ (Fig. 3.1). Iniciando-se de um vértice aleatório, a série temporal é construída através da realização deste caminho, na qual a probabilidade de mover-se do vértice n_i para o vértice n_j é dada por w_{ij} . Se for possível associar cada vértice da rede a um quantil

na série temporal resultante X , é possível construir a série temporal dividindo seu domínio em Q quantis e escolher aleatoriamente um valor pertencente ao quantil correspondente.

Na ausência de uma correspondência direta entre vértices e quantis assume-se suavidade na série temporal resultante. Desta forma, os vértices podem ser associados a quantis através do reordenamento da matriz de adjacência com pesos W , de tal forma que a mesma possua seus maiores elementos w_{ij} próximos da diagonal (SALESPARDO et al., 2007). Este reordenamento é feito utilizando-se a técnica do recozimento simulado (em inglês, simulated annealing) (KIRKPATRICK et al., 1983) e uma descrição do mesmo pode ser encontrada no Capítulo 7. A série temporal resultante será o mais suave possível — uma propriedade muito comum em diversas séries temporais empíricas.

Tal como o mapeamento direto \mathcal{M}_{QT} , o mapeamento inverso \mathcal{M}_{QT}^{-1} também possui algumas propriedades importantes. Ele é também sobrejetivo. Dada uma rede complexa g , o mapeamento irá produzir uma série temporal $X = \mathcal{M}_{QT}^{-1}[g, \epsilon]$ sobre uma realização ϵ , contudo, redes complexas distintas g e g' podem ser mapeadas sobre uma mesma série temporal $X = \mathcal{M}_{QT}^{-1}[g, \epsilon] = \mathcal{M}_{QT}^{-1}[g', \epsilon']$. Apesar deste mapeamento não ser estritamente um-para-um devido à presença de estocasticidade, ou seja, $\mathcal{M}_{QT}^{-1}[g, \epsilon] \neq \mathcal{M}_{QT}^{-1}[g, \epsilon']$, as séries temporais resultantes terão propriedades bastante similares. Em contraste, mapeamentos inversos anteriores (STROZZI et al., 2009; HARAGUCHI et al., 2009) são dependentes da escolha arbitrária na enumeração dos vértices, o que torna as séries temporais resultantes altamente sensíveis a enumeração escolhida.

Sumarizando, ao algoritmo a seguir descreve os passos necessários para a obtenção de uma rede complexa a partir de uma série temporal qualquer (mapeamento direto, \mathcal{M}_{QT}), como também, os passos necessários para a obtenção de uma série temporal com T pontos a partir de uma rede complexa qualquer (mapeamento inverso, \mathcal{M}_{QT}^{-1}).

Mapeamento direto \mathcal{M}_{QT} :

- I **Particionamento da série temporal.** Dada uma série temporal X com T pontos, divide-se X em Q quantis (conforme a Equação (3.1));
- II **Determinação dos vértices da rede associada.** Cada quantil q_i ($i = 1, 2, \dots, Q$) é associado a um vértice $n_i \in \mathcal{N}$ na rede g correspondente, nesta mesma ordem;
- III **Determinação das conexões entre vértices da rede.** Dois vértices n_i e n_j são conectados na rede com uma aresta $(n_i, n_j, w_{ij}) \in \mathcal{L}$, onde o peso w_{ij} de cada aresta é dado pelo número de vezes que um dado ponto x_t no quantil q_i é seguido por um ponto x_{t+1} no quantil q_j .

A rede resultante terá Q vértices e o peso total das arestas será dado por $T - 1$.

Mapeamento inverso \mathcal{M}_{QT}^{-1} :

- I **Determinação da matriz de transição W .** Dada uma rede complexa g com Q nós, primeiramente sua matriz de adjacência com pesos W é normalizada, de tal forma que, $\sum_j w_{ij} = 1$. Desta forma, cada elemento w_{ij} representa a probabilidade de transição do vértice n_i para o vértice n_j .
- II **Associação entre os vértices de g com os quantis em X .** A associação entre vértices e quantis é obtida através do reordenamento de W de modo que seus valores mais elevados estejam próximos da diagonal e a série temporal resultante seja o mais suave possível (SALES-PARDO et al., 2007).
- III **Determinação do domínio de X .** Assume-se que os valores de X estão uniformemente distribuídos no intervalo $[0, 1]$. Desta forma, este intervalo é dividido em Q partes iguais.
- IV **Construção dos elementos de X .** A série temporal X é construída movendo-se repetidamente do vértice n_i para o vértice n_j com probabilidade w_{ij} . A cada passo é assinalado aleatoriamente um número real pertencente ao quantil correspondente q_j , e este processo é repetido T vezes.

A série temporal resultante terá T pontos, que correspondem aos T passos do caminho aleatório sobre a rede complexa g .

4 RESULTADOS

Com o propósito de verificar quais propriedades de uma série temporal (rede complexa) são recuperadas quando \mathcal{M}_{QT} e \mathcal{M}_{QT}^{-1} são aplicados sequencialmente, introduz-se um conjunto de séries temporais que varia de periódica à totalmente aleatória:

$$x(t) = \begin{cases} \text{mod}(x(t-1) + \delta + \eta, 1), & \text{com probabilidade } p \\ \text{mod}(x(t-1) + \delta, 1), & \text{caso contrário} \end{cases} \quad (4.1)$$

onde η é uma variável aleatória obtida de uma distribuição uniforme entre $[0, 1]$ e p parametriza a probabilidade com que este ruído modifica a série periódica original. Tomando $p = 0, 10^{-2}, 10^{-1}, 10^0$ e $\delta = 0.05$, gera-se um conjunto de séries temporais com $T = 320$ pontos cada (Fig. 4.1).

O mapeamento direto \mathcal{M}_{QT} é então aplicado nas séries temporais obtidas através da relação (4.1) utilizando-se $Q = 20$ quantis e as redes complexas resultantes são obtidas. Estas séries temporais e redes complexas são referenciadas como “primeira geração” de séries temporais e redes complexas, respectivamente. A Figura 4.1 mostra que séries temporais com propriedades distintas são mapeadas em redes complexas com propriedades topológicas distintas. Especificamente, a medida que as séries temporais se tornam mais aleatórias as redes complexas associadas também se tornam mais aleatórias, como no modelo de redes pequeno mundo de Watts & Strogatz descrito na Seção 2.1.4.2 (WATTS; STROGATZ, 1998).

Em seguida, o mapeamento inverso \mathcal{M}_{QT}^{-1} é aplicado na primeira geração de redes complexas para se obter a segunda geração de séries temporais, cada uma com $T = 320$ pontos. Por simplicidade, assinala-se cada quantil ao mesmo quantil da primeira geração de séries temporais. É visível a similaridade entre a primeira geração de séries temporais X e segunda geração de séries temporais $\mathcal{M}_{QT}^{-1}[\mathcal{M}_{QT}[X], \epsilon]$, independente dos valores de p (Fig. 4.1).

Finalmente, o mapeamento proposto \mathcal{M}_{QT} é aplicado na segunda geração de séries temporais utilizando-se $Q = 20$ quantis para se obter a “segunda geração” de redes correspondentes. É notável que o mapeamento proposto é capaz de produzir a primeira geração de redes complexas g e a segunda geração de redes complexas

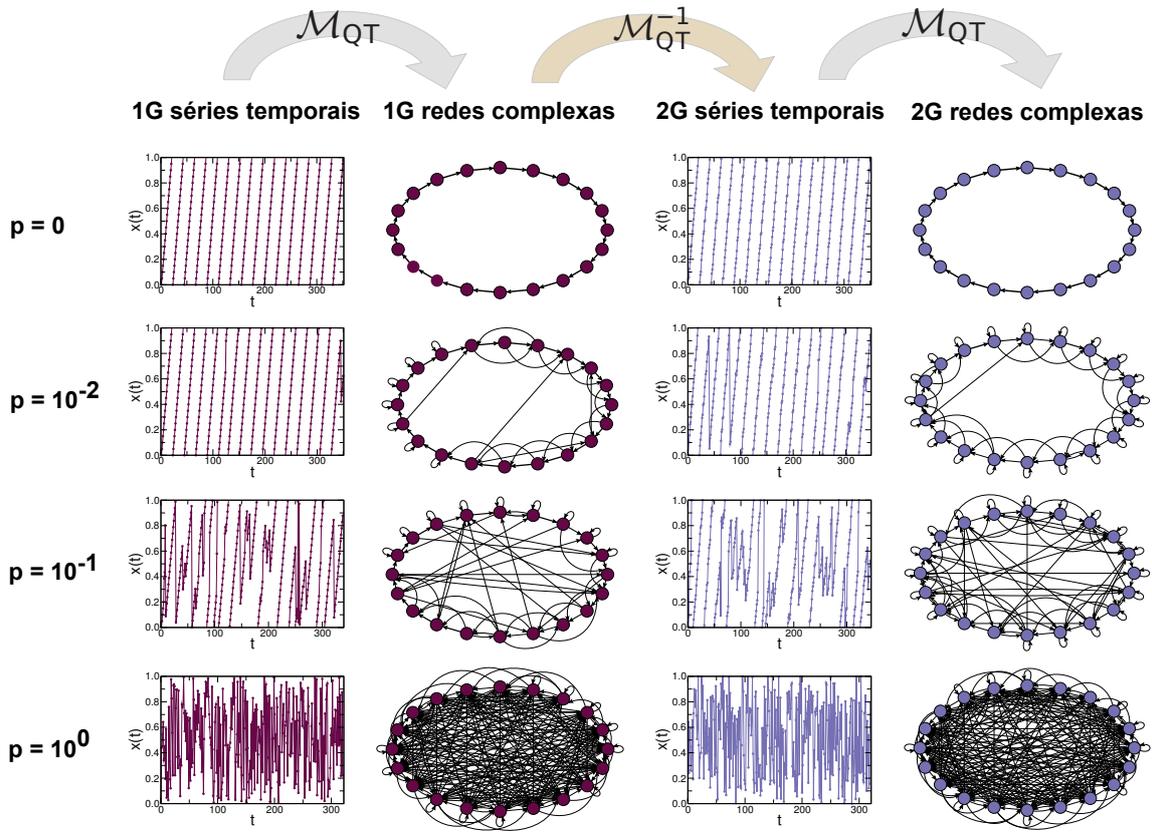


Figura 4.1 - Avaliação qualitativa da fidelidade do mapeamento proposto. Note que a primeira geração (**1G**) e a segunda geração (**2G**) de séries temporais e redes complexas possuem propriedades similares (Fig. 4.2-4.3), sugerindo-se que pode ser possível utilizar a análise de séries temporais para quantificar redes complexas, e vice-versa.

$\mathcal{M}_{QT}(\mathcal{M}_{QT}^{-1}(g))$ com topologias similares para todos os valores de p (Fig. 4.1).

Demonstra-se quantitativamente a confiabilidade do mapeamento proposto no domínio temporal \mathcal{T} através da comparação de diferentes propriedades estatísticas das primeira e segunda gerações de séries temporais. Em ambos os casos, as funções de autocorrelação e os espectros de potência são bastante similares e revelam a presença de periodicidade nas séries associadas quando $p = 0$, na qual desaparece com o incremento de p . Além disso, ambas as primeira e segunda gerações de séries temporais possuem valores uniformemente distribuídos entre 0.0 e 1.0, para todos os valores de p (Fig. 4.2).

A fidelidade do mapeamento no domínio de redes complexas \mathcal{G} é demonstrado atra-

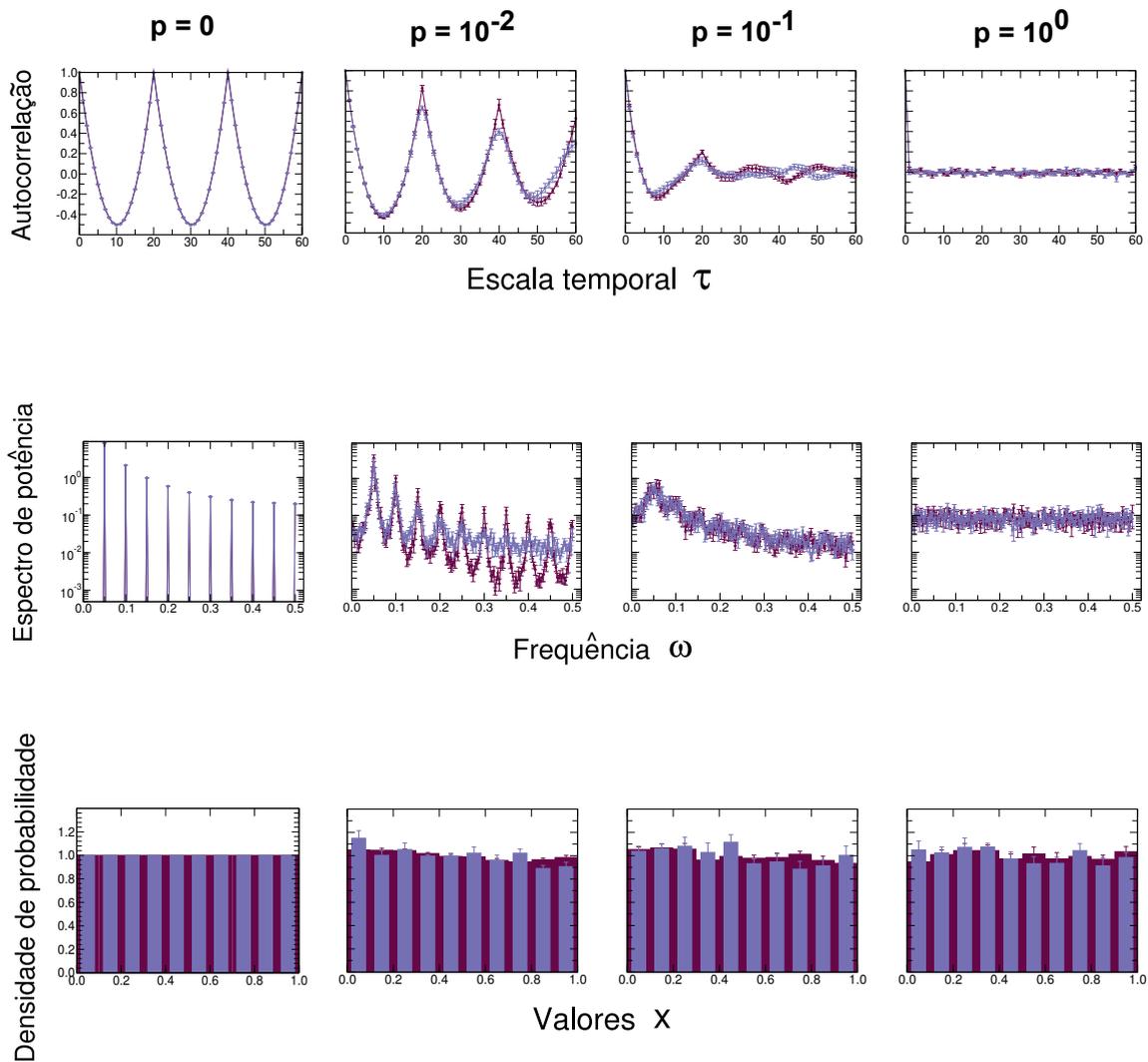


Figura 4.2 - Comparação de algumas propriedades estatísticas associadas a primeira geração (**1G**) e a segunda geração (**2G**) de séries temporais. São comparadas as médias destas propriedades sob 10 realizações distintas. As barras de erro denotam os desvios padrão sob tais realizações. Para ambas as gerações de séries temporais, as funções de autocorrelação, os espectros de potência e as distribuições de seus valores são bastante similares, independente dos valores de p .

vés da comparação de várias propriedades estatísticas das primeira e segunda gerações de redes complexas, incluindo a distribuição da força de entrada (em inglês, in-strength), o peso das arestas e o comprimento do menor caminho. A força de entrada é unitária em cada vértice quando as primeira e segunda gerações de redes complexas são regulares ($p = 0.0$) e quando tais redes se tornam aleatórias, a distribuição da força de entrada se espalha devido a redistribuição dos pesos. Note

que a força de saída (em inglês, out-strength) de cada vértice é unitária em todos os casos, já que os pesos são probabilidades markovianas. Os pesos das arestas são unitários quando as primeira e segunda gerações de redes complexas são regulares ($p = 0.0$). Quando p aumenta, a distribuição dos pesos das arestas mostra a presença de pequenos pesos ($w \sim 0.0$) como também grandes pesos ($w \sim 1$). O comprimento do menor caminho é calculado como a soma mínima do inverso dos pesos sob um caminho entre dois vértices. Os comprimentos dos menores caminhos são uniformemente distribuídos quando as primeira e segunda gerações de redes complexas são regulares ($p = 0.0$). Quando p aumenta, os atalhos aleatórios diminuem as distâncias entre os vértices, embora em alguns casos, comprimentos do caminho mais longos surgem devido a redistribuição dos pesos. À medida que as redes se tornam mais aleatórias, as distribuições dos menores caminhos assumem um pico (Fig. 4.3).

Pode-se observar que as propriedades topológicas da primeira geração de redes complexas são recuperadas na segunda geração de redes complexas para todos os valores de p . De modo geral, os resultados aqui apresentados (Fig. 4.1-4.3) indicam que o método proposto nesta tese é capaz de preservar informações estruturadas e não estruturadas nos domínios série temporal e rede complexa, sob sucessivas aplicações dos mapeamentos direto e inverso.

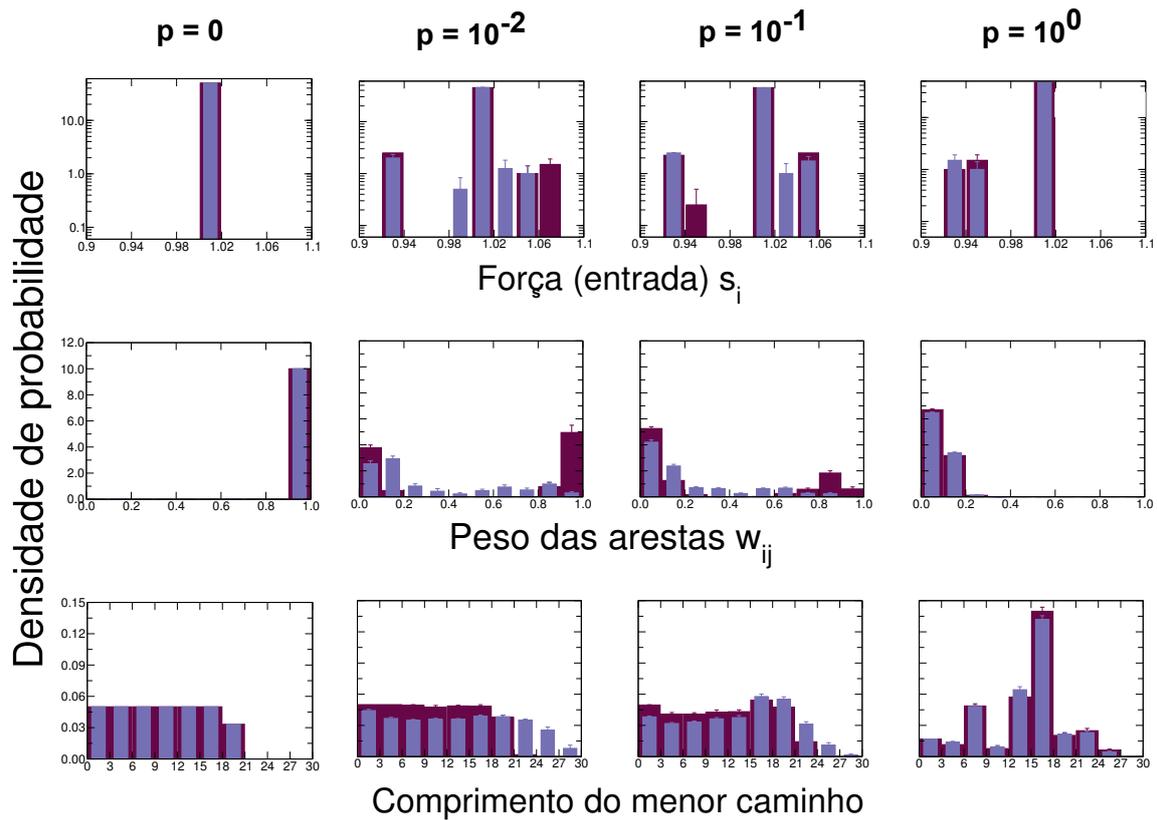


Figura 4.3 - Comparação de algumas propriedades topológicas associadas a primeira geração (**1G**) e a segunda geração (**2G**) de redes complexas. São comparados os valores médios de tais propriedades sob 10 realizações diferentes. As barras de erro denotam os desvios padrão sob tais realizações. Para ambas as gerações de redes complexas, as distribuições das forças de entrada, os pesos das arestas e os comprimentos dos menores caminhos são bastante similares, para todos os valores de p .

4.1 Aplicação do mapeamento direto em séries temporais caóticas

Com o intuito de enfatizar o potencial do mapeamento direto, o mesmo é aplicado em duas séries temporais pertencentes a sistemas dinâmicos distintos. A primeira série temporal é a variável $x(t)$ do sistema caótico de Lorenz (2.14), com $\sigma = 10.0$, $r = 28.0$, e $b = 8/3$. As soluções numéricas destas equações se desenvolvem em um atrator imerso no espaço tridimensional com coordenadas (x, y, z) (LORENZ, 1963). A trajetória do sistema rotaciona ao redor de um dos pontos de equilíbrio instáveis e eventualmente escapa para a órbita do outro ponto de equilíbrio. Este comportamento é reconhecido em sua variável $x(t)$ já que seus valores oscilam entre valores positivos e negativos (Fig. 4.4).

A segunda série temporal é a variável $x(t)$ do sistema caótico de Rossler (2.26), com $a = 0.432$, $b = 2.0$, e $c = 4.0$. Seu espaço de fase gera um atrator caótico com um único lóbulo, em contraste com o atrator de Lorenz que possui dois. A trajetória dentro do atrator segue uma espiral próxima do plano (x, y) , ao redor de um ponto de equilíbrio instável. Uma vez que a trajetória rotaciona por várias vezes, o segundo ponto de equilíbrio a influencia, ocasionando um salto na dimensão z (ROSSLER, 1976). Este comportamento gera um padrão oscilatório quasiperiódico na variável $x(t)$ (Fig. 4.4).

Em ambos os casos, o mapeamento direto \mathcal{M}_{QT} é aplicado utilizando-se $T = 10,000$ pontos da variável $x(t)$ das equações de Lorenz e Rossler e redes complexas são construídas utilizando-se $Q = 50$ quantis. Cada vértice é colorido de acordo com o módulo no qual está inserido. As redes complexas resultantes (Fig. 4.4) apresentam diferenças nítidas em suas topologias. A rede complexa do sistema de Lorenz apresenta uma estrutura densa com dois grandes módulos. Tal rede possui modularidade $M_o = 0.4547$, que é muito maior que a modularidade média $\langle M_o \rangle_{\mathcal{D}} = 0.0298 \pm 0.0018$, obtida de redes construídas através da aleatorização da série temporal original. Além disso, os dois lóbulos do atrator de Lorenz são mapeados nos dois maiores módulos na rede. Por outro lado, a rede complexa associada ao sistema de Rossler apresenta uma estrutura alongada devido a periodicidade presente na série temporal correspondente. Tal rede possui modularidade $M_o = 0.6437$, que é muito maior que a modularidade média $\langle M_o \rangle_{\mathcal{D}} = 0.0280 \pm 0.0017$, obtida de redes construídas a partir da aleatorização da série original. Além disso, esta rede possui cinco pequenos módulos em virtude das diferentes amplitudes geradas pelo atrator.

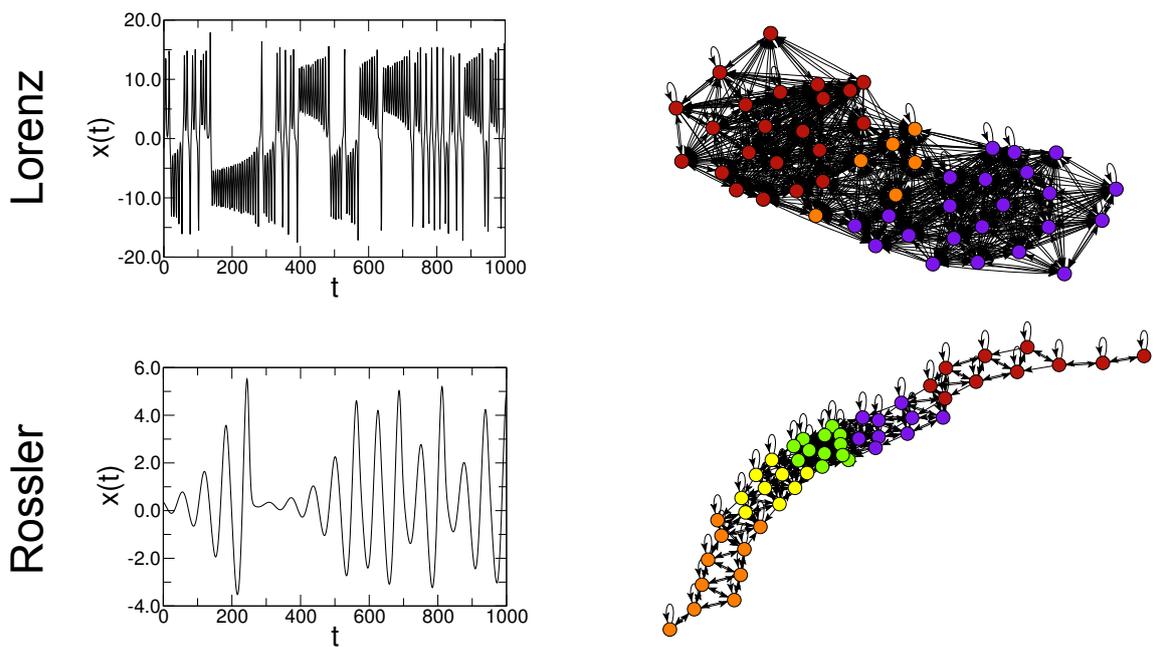


Figura 4.4 - Ilustração do mapeamento direto M_{QT} em séries temporais caóticas. São utilizados $T = 10,000$ pontos das variáveis $x(t)$ das equações de Lorenz e de Rossler e construídas redes complexas com $Q = 50$ quantis. As redes complexas resultantes apresentam diferenças nítidas em suas topologias. A rede associada ao sistema de Lorenz apresenta uma estrutura densa, onde os dois lóbulos do atrator são mapeados nos dois maiores módulos da rede complexa. Por outro lado, a rede complexa associada ao sistema de Rossler apresenta uma estrutura alongada decorrente do padrão quasiperiódico presente na série temporal correspondente.

4.2 Aplicação do mapeamento direto em séries temporais auto-similares

O mapeamento direto é também aplicado em um conjunto de séries temporais auto-similares. Uma série temporal é auto-similar se o espectro de potência correspondente, $P(\omega)$, é caracterizado por uma lei de potência (MALAMUD; TURCOTTE, 1999), ou seja:

$$P(\omega) \sim \omega^{-\beta}, \quad (4.2)$$

onde o expoente β é uma medida da intensidade da persistência (correlação) na série temporal associada. Mais especificamente, valores estritamente positivos de β indicam persistência e quanto maior o seu valor, maior a persistência. O valor $\beta = 0$ indica a ausência de persistência, e por fim, valores estritamente negativos de β indicam anti-persistência e quanto menor o seu valor, maior a anti-persistência.

A Figura 4.5 apresenta exemplos de séries temporais auto-similares com $T = 1,000$, geradas a partir de valores de β iguais a -1.0 ; -0.5 ; 0.0 ; 0.5 ; 1.0 e 2.0 . À medida que o valor de β aumenta as séries temporais correspondentes se tornam mais persistentes, ou seja, seus valores adjacentes se tornam mais correlacionados. A Figura 4.6 apresenta os espectros associados as séries da Figura 4.5. Note que tais espectros são caracterizados por uma lei de potência com inclinação dada por $-\beta$ (em escala log-log).

O mapeamento direto \mathcal{M}_{QT} é aplicado utilizando-se séries temporais com $T = 10,000$ pontos e $Q = 50$ quantis (Fig. 4.5). As redes complexas resultantes apresentam diferenças significantes em topologia, conforme observado nas matrizes de adjacência com pesos associadas (Fig. 4.7). À medida que o valor de β aumenta, maior é a probabilidade dos valores adjacentes de uma série temporal estarem correlacionados, e desta forma, maior a probabilidade de vértices vizinhos estarem conectados entre si.

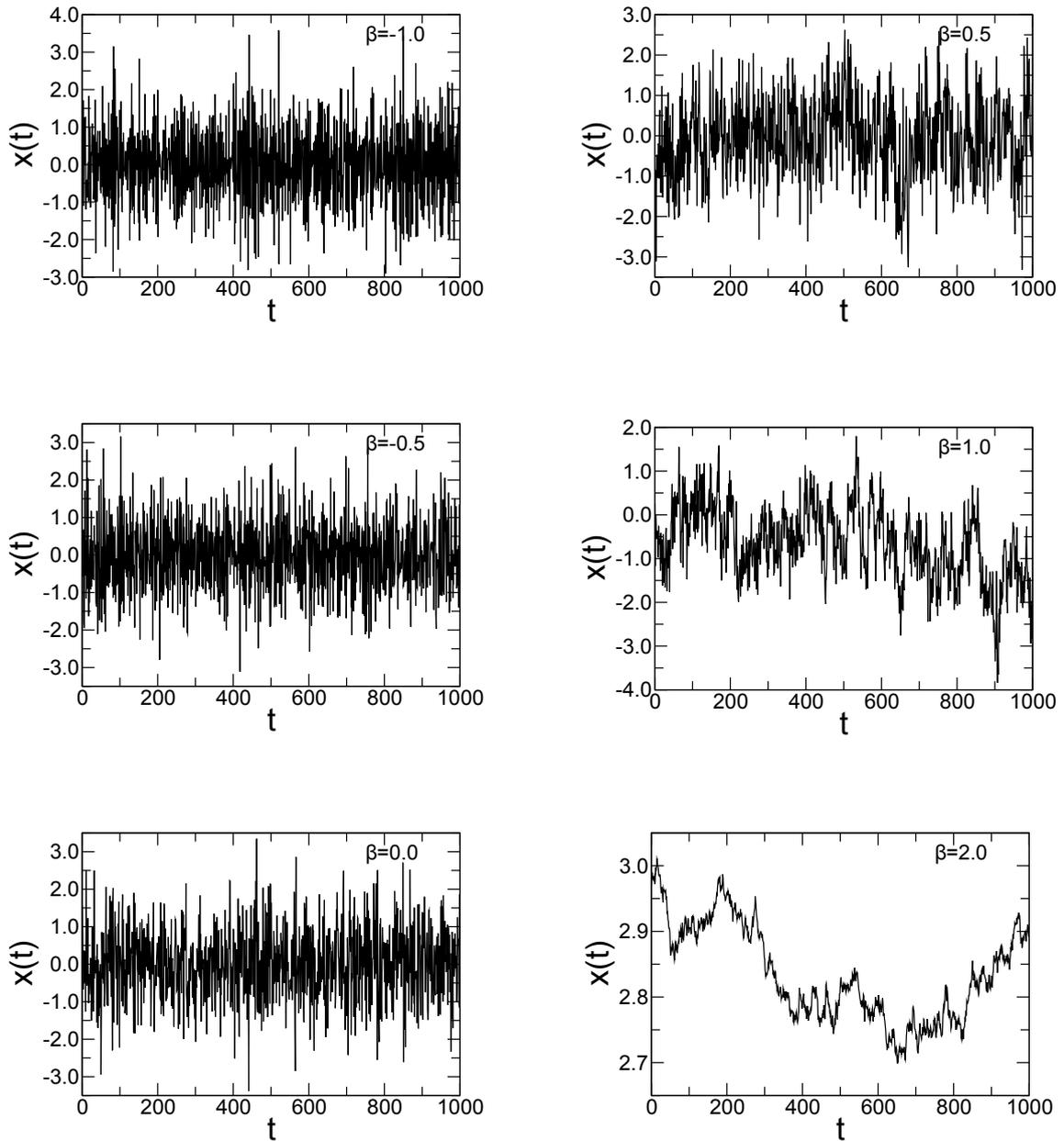


Figura 4.5 - Exemplos de séries temporais auto-similares com $-1.0 \leq \beta \leq 2.0$. Note que a medida que o valor de β aumenta as séries temporais correspondentes se tornam mais persistentes, ou seja, seus valores adjacentes se tornam mais correlacionados.

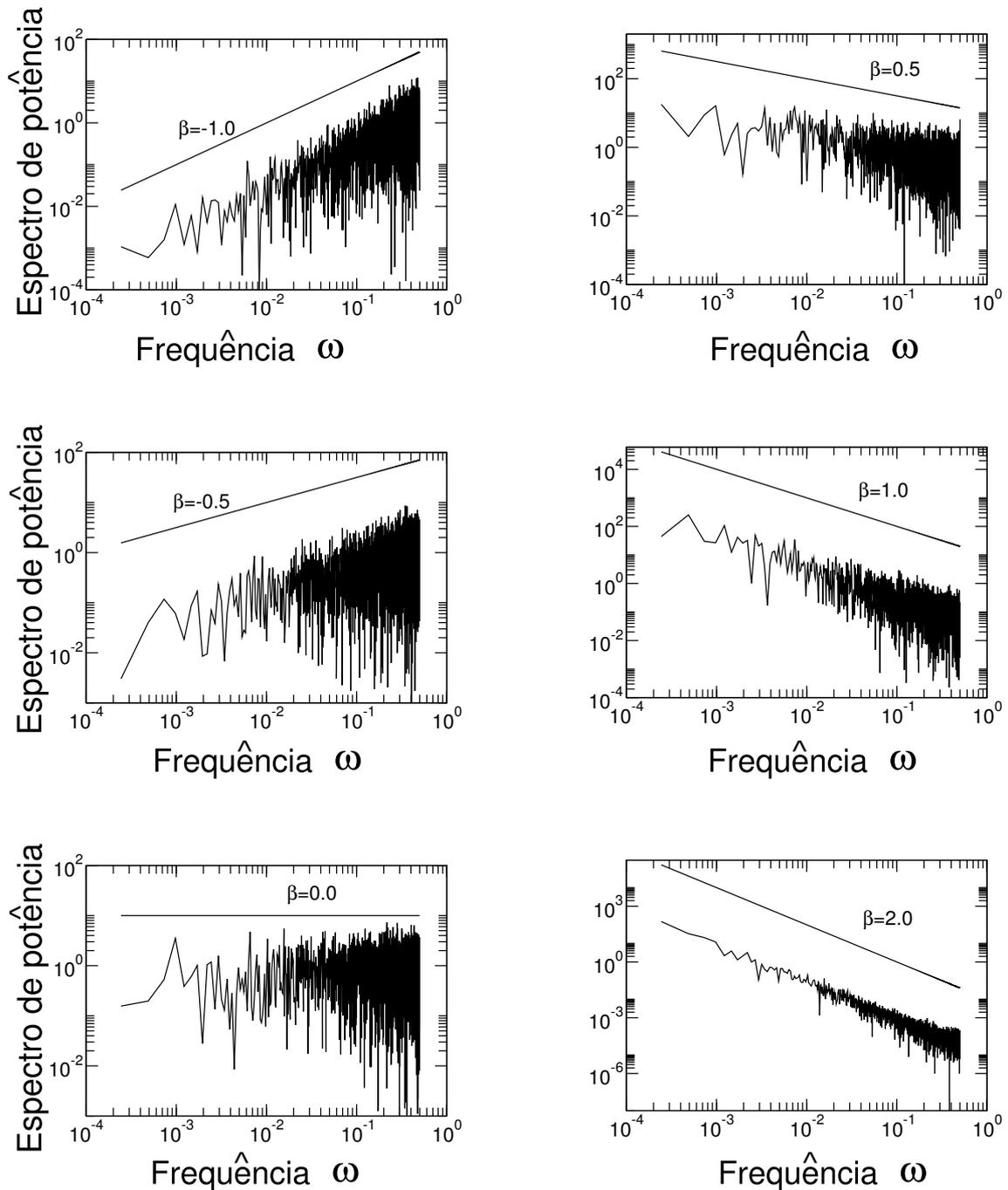


Figura 4.6 - Espectros de potência (em escala log-log) associados as séries temporais auto-similares da Figura 4.5. Em todos os casos, os espectros de potência são caracterizados por uma lei de potência com inclinação dada por $-\beta$.

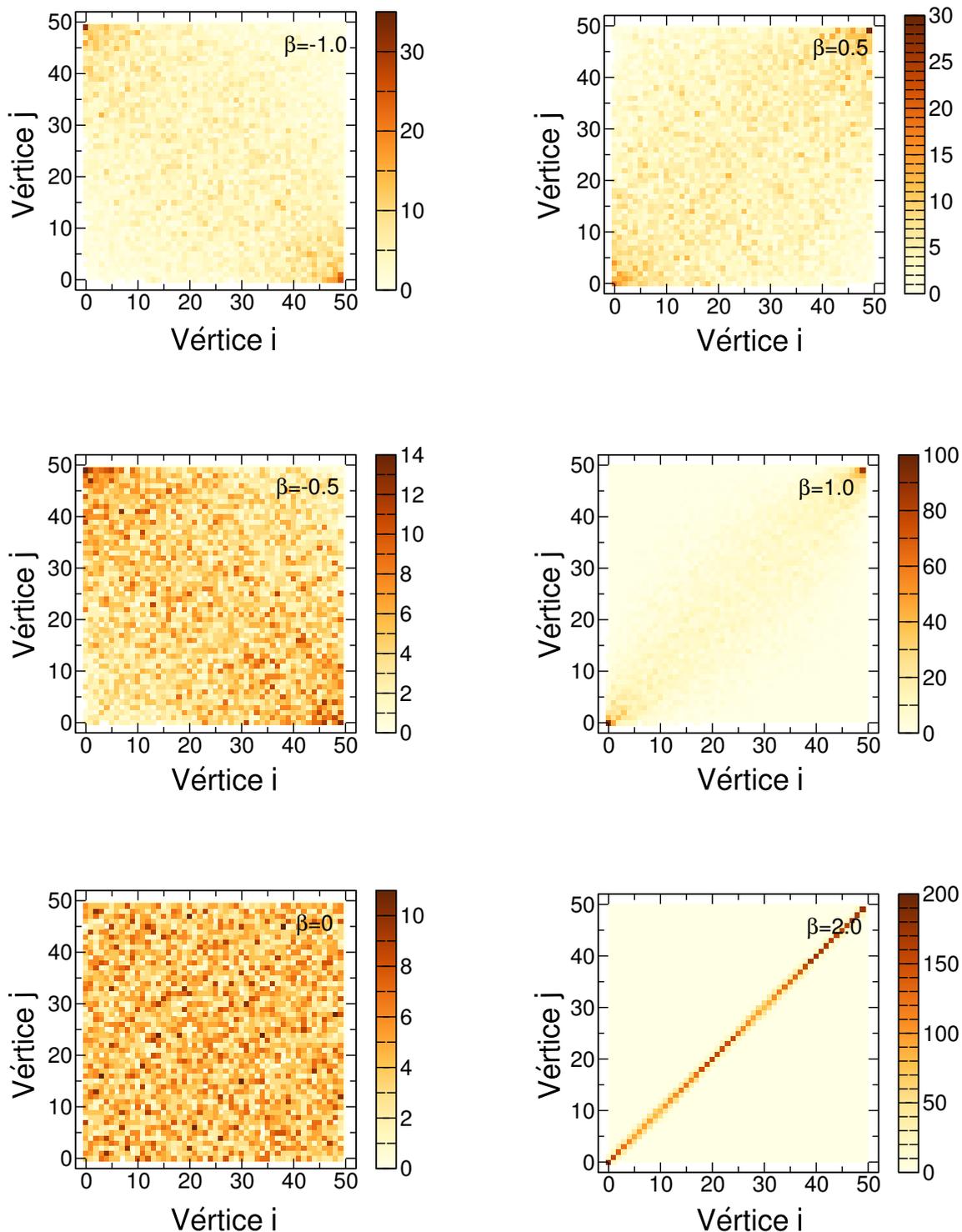


Figura 4.7 - Matrizes de adjacência com pesos obtidas através da aplicação do mapeamento direto \mathcal{M}_{QR} nas séries temporais auto-similares da Figura 4.5. Quanto maior a anti-persistência na série temporal, maior é a probabilidade de vértices distantes estarem conectados. A medida que β aumenta, a correlação na série associada também aumenta e maior é a probabilidade de vértices vizinhos estarem conectados na rede correspondente.

As diferenças em topologia observadas na Figura 4.7 ficam mais evidentes em gráficos do coeficiente de agrupamento C e do comprimento do menor caminho médio L em função do coeficiente β (Fig. 4.8). Os valores de C e L foram obtidos com base em redes sem peso, obtidas a partir da aplicação da operação de limiarização nas redes com peso da Figura 4.7. Como esperado, um ruído branco ($\beta = 0.0$) é mapeado em uma rede completamente aleatória. Além disso, as correlações crescentes (ou maior persistência) produzem redes cada vez mais regulares, onde o fenômeno de pequeno mundo tende a desaparecer. Em conjunto, estes resultados atestam que o mapeamento proposto traduz de modo eficaz propriedades puramente geométricas de uma série temporal (auto-similaridade, por exemplo) em características topológicas bem definidas na rede correspondente.

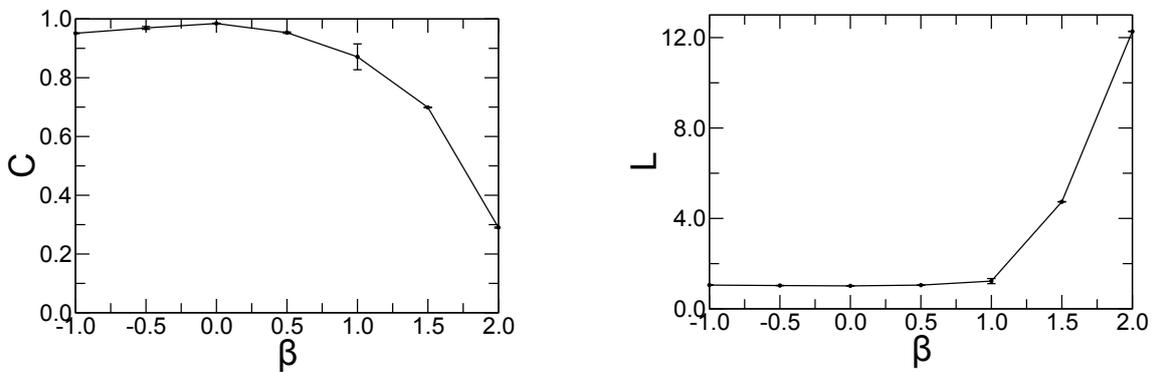


Figura 4.8 - Coeficiente de agrupamento e comprimento do caminho médio em função do coeficiente β associados as redes da Figura 4.7. São comparadas as médias destas propriedades sob 10 realizações distintas. As barras de erro denotam os desvios padrão sob tais realizações. Como esperado, um ruído branco ($\beta = 0.0$) gera uma rede completamente aleatória. Observe que séries com correlações crescentes (ou maior persistência) produzem redes cada vez mais regulares, onde o fenômeno de pequeno mundo tende a desaparecer.

4.3 Aplicação do mapeamento direto em séries temporais de batimento cardíaco

Com o intuito de ilustrar o potencial do mapeamento proposto em aplicações reais, o mesmo é aplicado ao problema de detecção de diferenças em batimentos cardíacos em séries temporais de pacientes sadios e doentes (GOLDBERGER et al., 2000). Especificamente, utiliza-se duas séries temporais de batimento cardíaco do PhysioNet (PHYSIONET...); uma de um paciente sadio e outra de um paciente com falência cardíaca (Fig. 4.9). A série temporal do paciente sadio é não-estacionária e irregular. Por outro lado, a série temporal de um paciente com falência cardíaca está associada a emergência de regularidade excessiva. O mapeamento direto \mathcal{M}_{QT} é aplicado em séries temporais normalizadas de 100-minutos de pacientes sadios e com falência cardíaca amostradas a 0.01 segundos ($T = 10,000$ pontos) e $Q = 50$ quantis (Fig. 4.9). As redes complexas resultantes apresentam diferenças nítidas em topologia, especialmente aparente na rede associada ao paciente não-sadio. Esta rede possui um “cluster” em separado decorrente da presença de intermitências (em inglês, bursts) na série temporal correspondente.

A robustez dos resultados encontrados na Figura 4.9 é demonstrada aplicando o mapeamento direto \mathcal{M}_{QT} nas séries estudadas utilizando-se diferentes valores de Q . A Figura 4.10 sugere que o mapeamento direto é capaz de produzir redes complexas com topologias similares as encontradas na Figura 4.9, independente dos valores de Q . Como uma outra demonstração de robustez, aplica-se o mapeamento direto \mathcal{M}_{QT} em séries temporais de batimento cardíaco de diferentes pacientes sadios e não-sadios. A Figura 4.11 sugere que o mapeamento direto é capaz de produzir redes complexas com topologias similares as encontradas na Figura 4.9, para pacientes sadios e não-sadios.

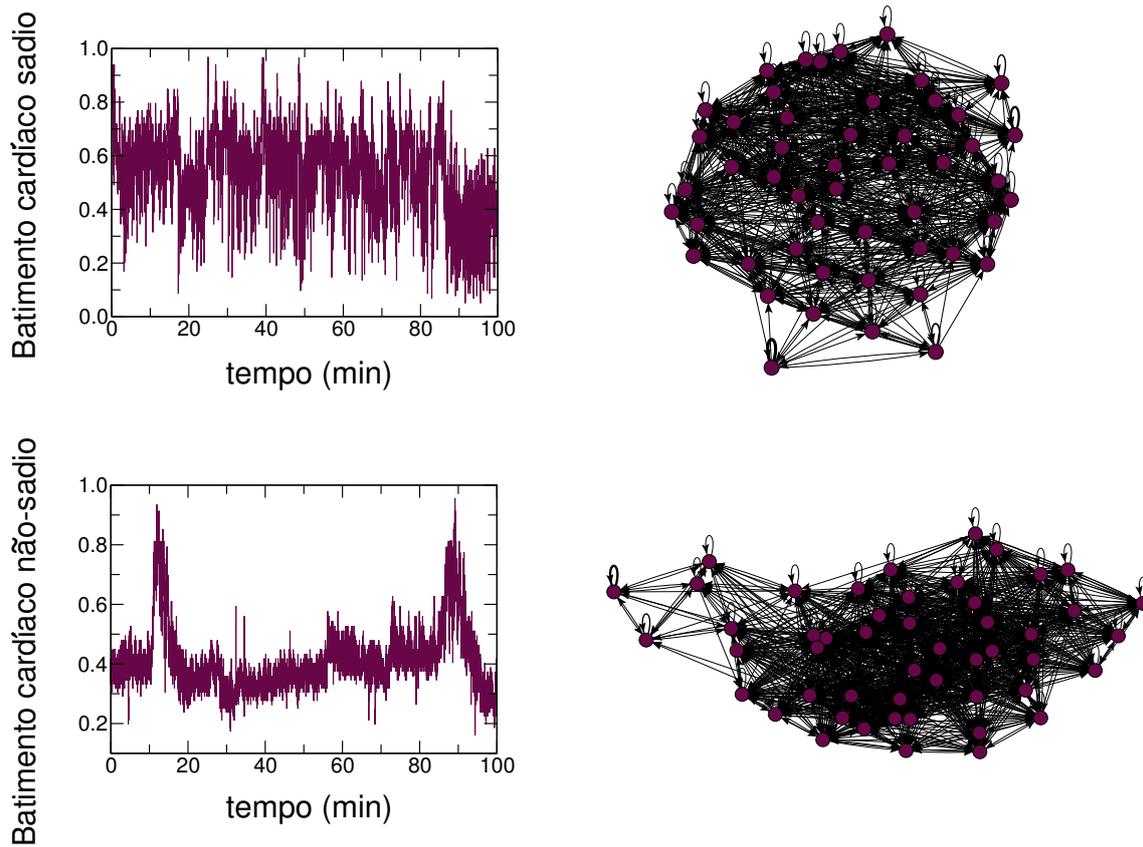


Figura 4.9 - Ilustração do mapeamento direto \mathcal{M}_{QT} no problema de detecção de diferenças na dinâmica das séries temporais associadas a pacientes em diferentes condições de saúde. São utilizadas séries temporais normalizadas de batimento cardíaco a 100-minutos de pacientes sadios e com falência cardíaca. Redes complexas são construídas utilizando-se $Q = 50$ quantis. As redes resultantes apresentam diferenças topológicas, com a presença de um cluster em separado na rede associada ao paciente não-sadio.

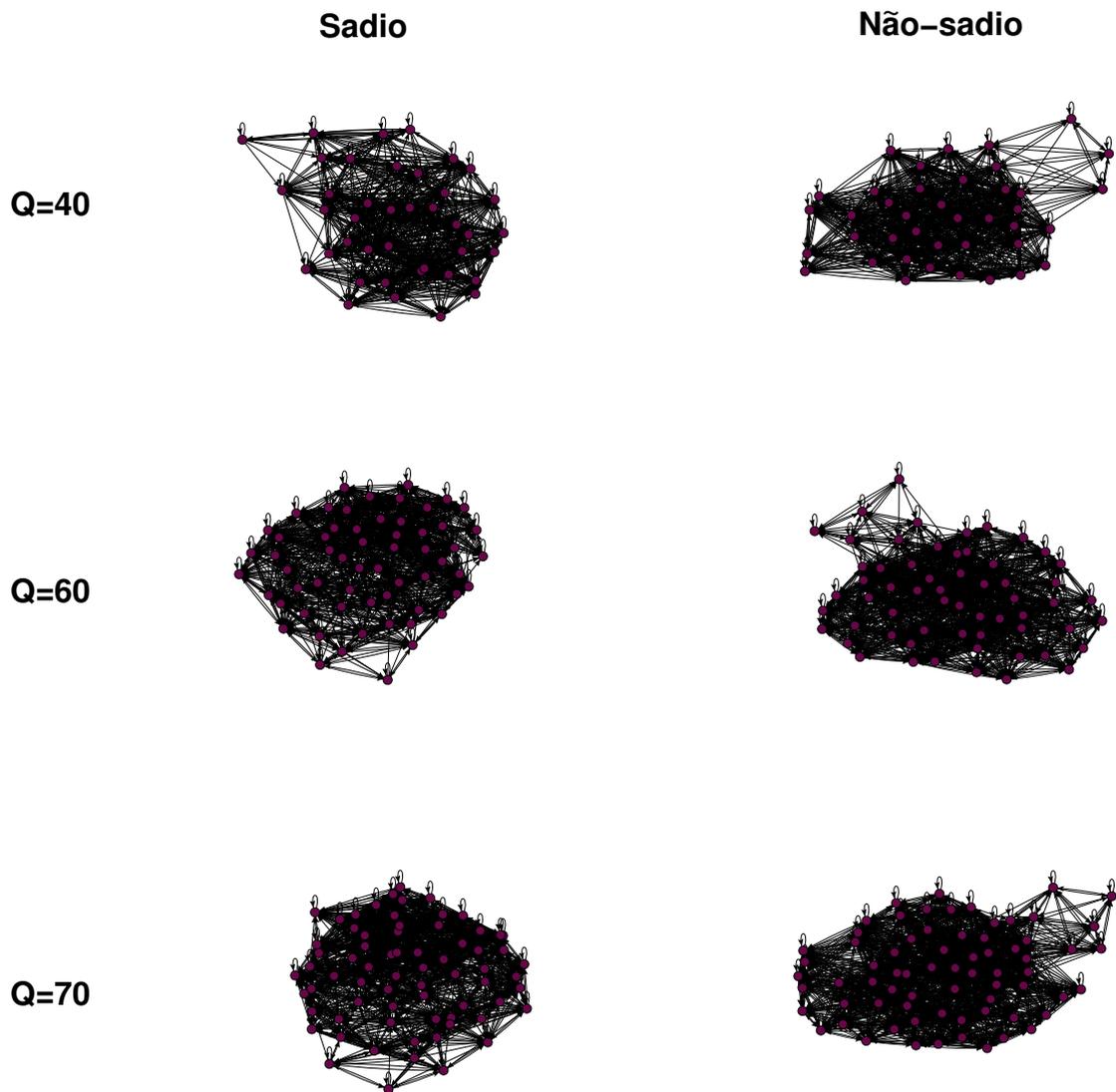
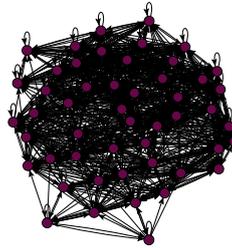
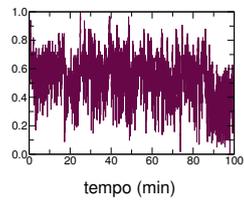
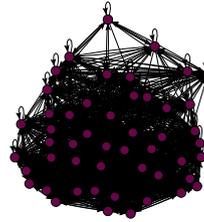
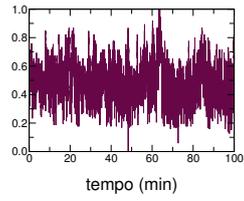
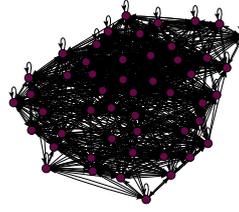
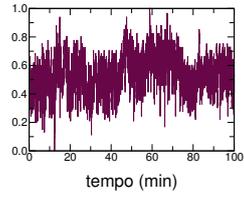


Figura 4.10 - Três aplicações do mapeamento \mathcal{M}_{QT} utilizando-se $Q = 40, 60$ e 70 quantis em séries temporais de pacientes sadios e não-sadios. Note a similaridade entre tais redes com as redes apresentadas na Figura 4.9, atestando a robustez dos resultados associados ao mapeamento direto, independente dos valores de Q .

Batimento cardíaco sadio



Batimento cardíaco não-sadio

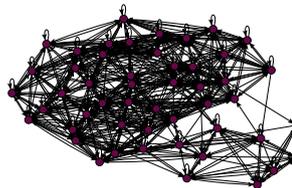
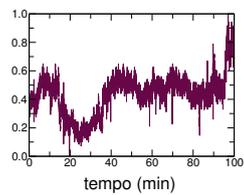
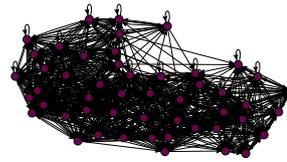
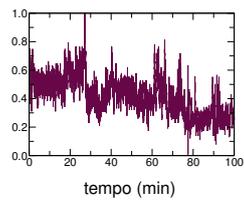
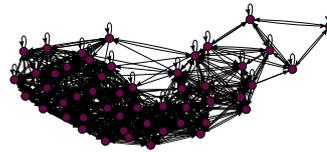
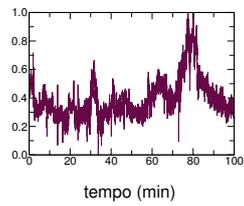


Figura 4.11 - Três aplicações do mapeamento \mathcal{M}_{QT} utilizando-se $Q = 50$ quantis em diferentes séries temporais associadas a pacientes sadios e não-sadios. Independente do número de pacientes, as redes complexas resultantes são visualmente similares as apresentadas na Figura 4.9.

4.4 Aplicação do mapeamento inverso em redes complexas reais

O potencial do mapeamento inverso em aplicações reais é ilustrado aplicando-o em duas redes complexas pertencentes a classes distintas. A primeira rede é a rede metabólica de *Arabidopsis thaliana*, que apresenta elevada modularidade e é caracterizada por uma estrutura em cadeias e um conjunto de hubs que são diretamente atingíveis entre si. A segunda rede é a rede de Internet nos Estados Unidos em 1997, que possui uma estrutura em estrela com diversos hubs e baixa modularidade, conforme mostra a Figura 4.12 (GUIMERÀ et al., 2007; NETGEO...).

Primeiramente, vértices são associados a quantis através do reordenamento da matriz de adjacência com pesos correspondente (SALES-PARDO et al., 2007) (Fig. 4.13) e séries temporais com $T = 100,000$ pontos cada são geradas através da aplicação do mapeamento inverso \mathcal{M}_{QT}^{-1} (Fig. 4.13). As séries temporais resultantes apresentam diferenças nítidas em dinâmica. Na primeira aplicação, as características topológicas da rede metabólica são traduzidas em séries temporais com elevado grau de persistência (ou correlações de longo alcance), devido a presença de cadeias em sua estrutura original. Por outro lado, cada vez que o caminhante aleatório atinge um dos diversos hubs na rede de Internet, existe uma probabilidade elevada do mesmo ser levado a diferentes partes da rede. Este comportamento produz um sinal caracterizado por uma série temporal de baixa persistência (correlações de curto alcance).

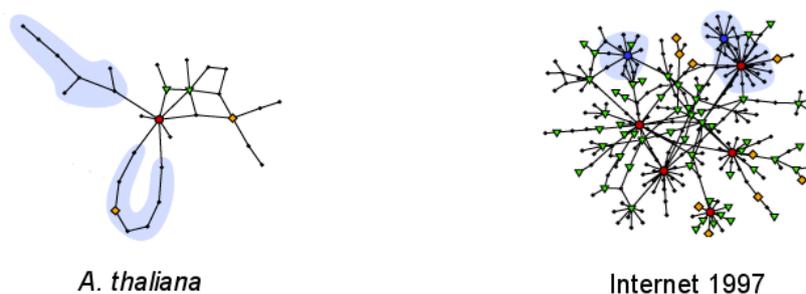


Figura 4.12 - Representações esquemáticas de duas redes complexas pertencentes a classes distintas. A primeira é a rede metabólica de *Arabidopsis thaliana*, que é caracterizada por uma estrutura em cadeias e um conjunto de hubs que são diretamente atingíveis entre si. A segunda é a rede de Internet nos Estados Unidos em 1997, que possui uma estrutura em estrela com diversos hubs. Fonte: Figura adaptada de Guimerà et al. (2007).

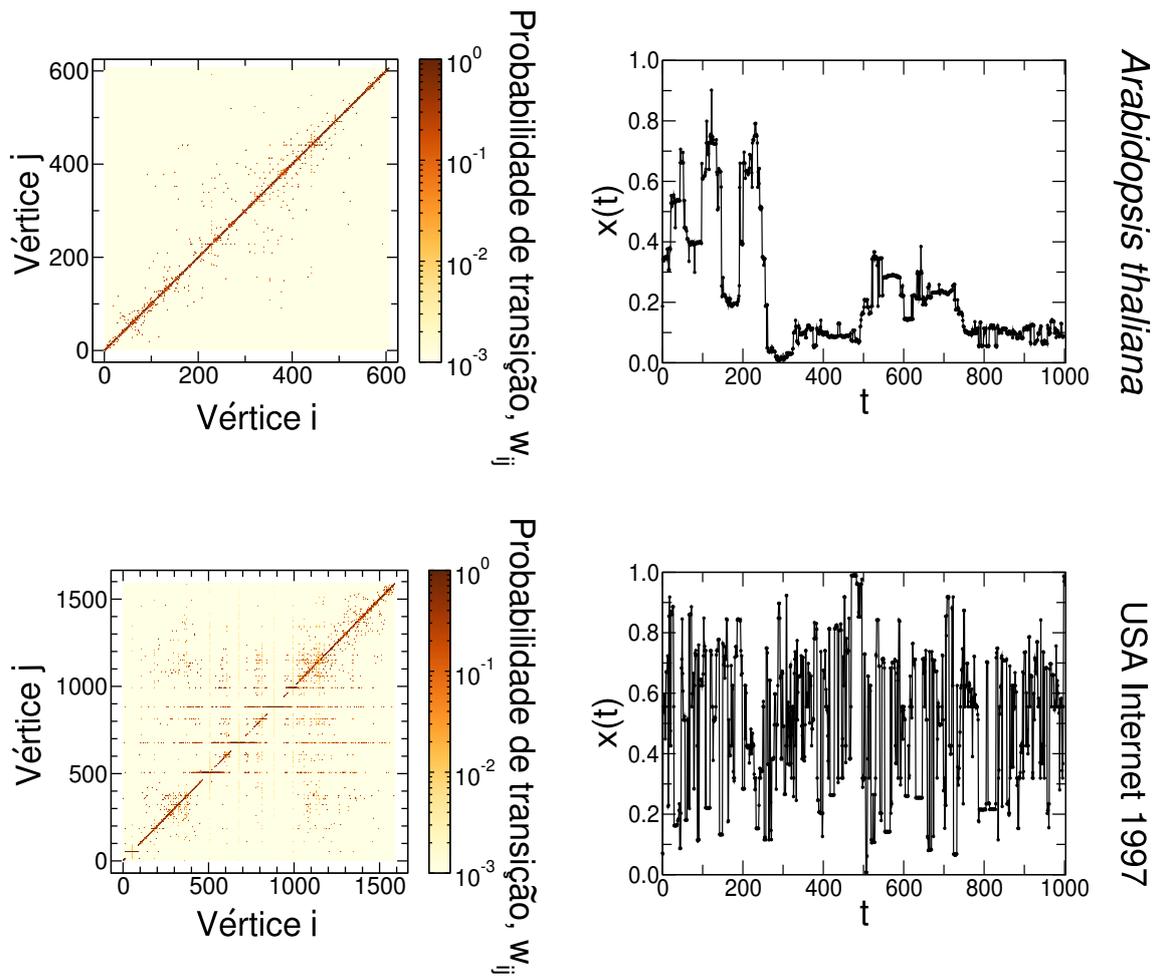


Figura 4.13 - Ilustração do mapeamento inverso em redes complexas reais de diferentes tipos — a rede metabólica *Arabidopsis thaliana* e a rede de Internet nos Estados Unidos em 1997. Primeiramente, vértices são associados a quantis através do reordenamento da matriz de adjacência com pesos correspondente (ambas em escala log-log). Em seguida, séries temporais são obtidas com $T = 100,000$ pontos cada utilizando-se redes com $Q = 607$ e $1,589$ vértices, respectivamente. As séries temporais resultantes apresentam diferenças nítidas em dinâmica.

As diferenças em dinâmica nas séries temporais resultantes são demonstradas quantitativamente através do cálculo da função de autocorrelação, do espectro de potência e do coeficiente de Hurst associados (Fig. 4.14). Note que as correlações de longo alcance presentes na rede metabólica são bem capturadas pela função de autocorrelação, pelo espectro de potência com inclinação $\beta = -1.4541$ e pelo valor médio do coeficiente de Hurst, dado por $H_u = 0.8579 \pm 0.0006$. Por outro lado, os resultados associados a rede de Internet mostram a assinatura de baixa correlação gerada por esta rede. Note que neste caso, o espectro do potência associado possui uma inclinação inferior, dada por $\beta = -1.1865$ e valor médio do coeficiente de Hurst também inferior, dado por $H_u = 0.7045 \pm 0.0014$.

A robustez destes mesmos resultados é demonstrada aplicando-se o mapeamento inverso \mathcal{M}_{QT}^{-1} , sob diferentes realizações, na rede metabólica *Arabidopsis thaliana* e na rede de Internet. Séries temporais com $T = 100,000$ pontos cada foram geradas através da aplicação sucessiva do mapeamento inverso \mathcal{M}_{QT}^{-1} em suas matrizes de adjacência com pesos reordenadas (Fig. 4.13). A Figura 4.15 sugere que dada uma rede complexa, o mapeamento inverso é capaz de produzir séries temporais com dinâmicas similares, sob diferentes realizações. De modo geral, os resultados apresentados demonstram que redes complexas com topologias distintas resultam em séries temporais com dinâmicas distintas.

A Figura 4.16 apresenta as matrizes de adjacência com pesos de redes pertencentes a diversas classes, incluindo redes metabólicas, de transporte aéreo e de interações entre proteínas. Primeiramente, vértices são associados a quantis através do reordenamento da matriz de adjacência com pesos correspondente e séries temporais com $T = 100,000$ pontos cada são geradas através da aplicação do mapeamento inverso \mathcal{M}_{QT}^{-1} . As séries temporais resultantes apresentam dinâmicas distintas.

A Tabela 4.1 apresenta algumas características das redes em estudo, tais como número de vértices, número de arestas e valores médios de modularidade associados as redes originais (M_o) e de suas versões aleatorizadas ($\langle M_o \rangle_{\mathcal{D}}$). Algumas características das séries temporais associadas são também apresentadas, tais como o valor médio de H_u e a inclinação média de seus espectros de potência. Apesar desta pequena amostra não permitir que sejam tiradas conclusões definitivas sobre a existência de correlações entre as propriedades de uma rede real e as características da série temporal correspondente, alguns resultados na Tabela 4.1, como os valores médios da modularidade e do expoente de Hurst para as redes metabólicas ou a Internet,

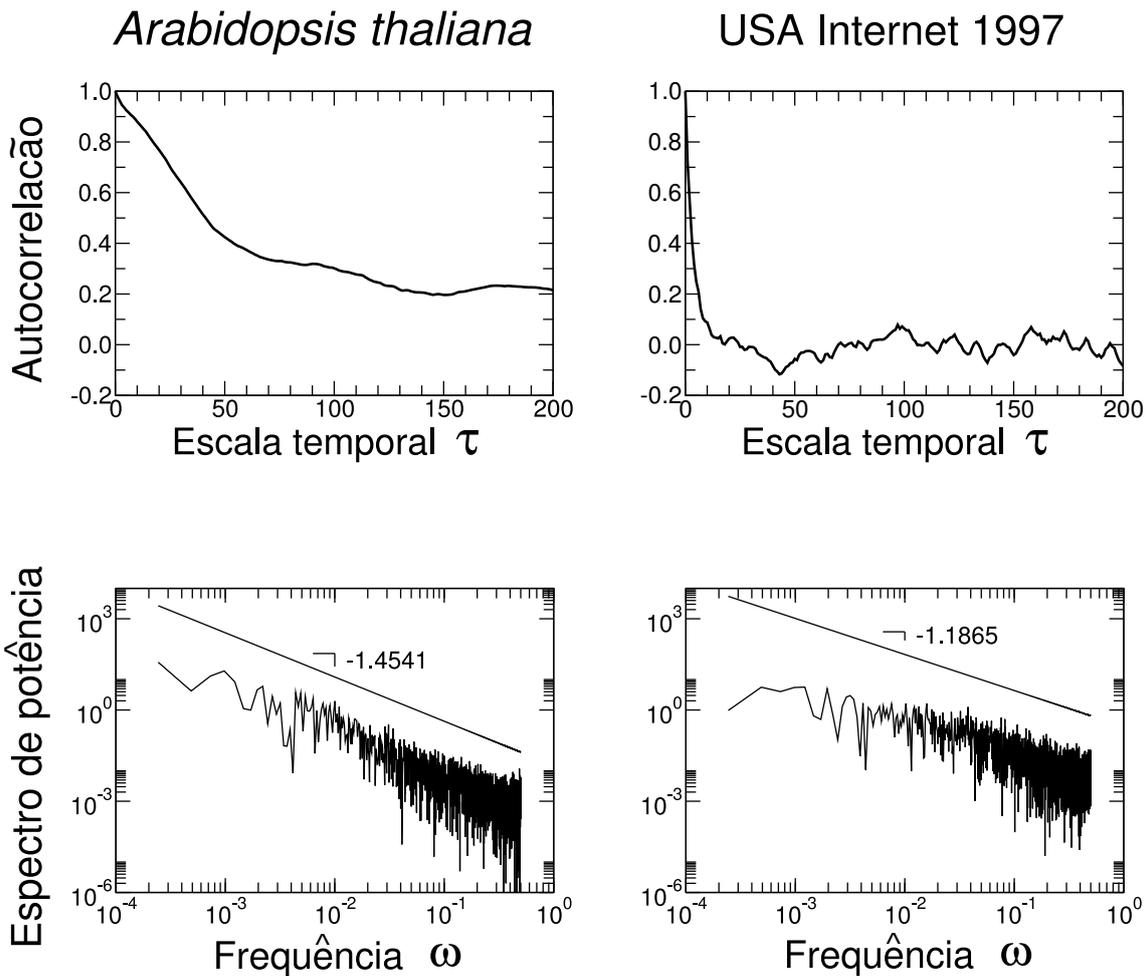


Figura 4.14 - Propriedades estatísticas das séries temporais apresentadas na Figura 4.13, obtidas da rede metabólica *Arabidopsis thaliana* e da rede de Internet nos Estados Unidos em 1997. As correlações de longo alcance presentes na rede metabólica são bem capturadas pela função de autocorrelação, pelo espectro de potência com inclinação $\beta = -1.4541$ e pelo valor médio do coeficiente de Hurst, dado por $H_u = 0.8579 \pm 0.0006$. Por outro lado, os resultados na rede de Internet mostram a assinatura de baixa correlação gerada por esta rede. Neste caso, o espectro do potência associado possui uma inclinação inferior, dada por $\beta = -1.1865$ e um coeficiente de Hurst também inferior, dado por $H_u = 0.7045 \pm 0.0014$.

sugerem fortemente que estudos mais aprofundados devem ser realizados sobre este tópico, como prosseguimento desta tese.

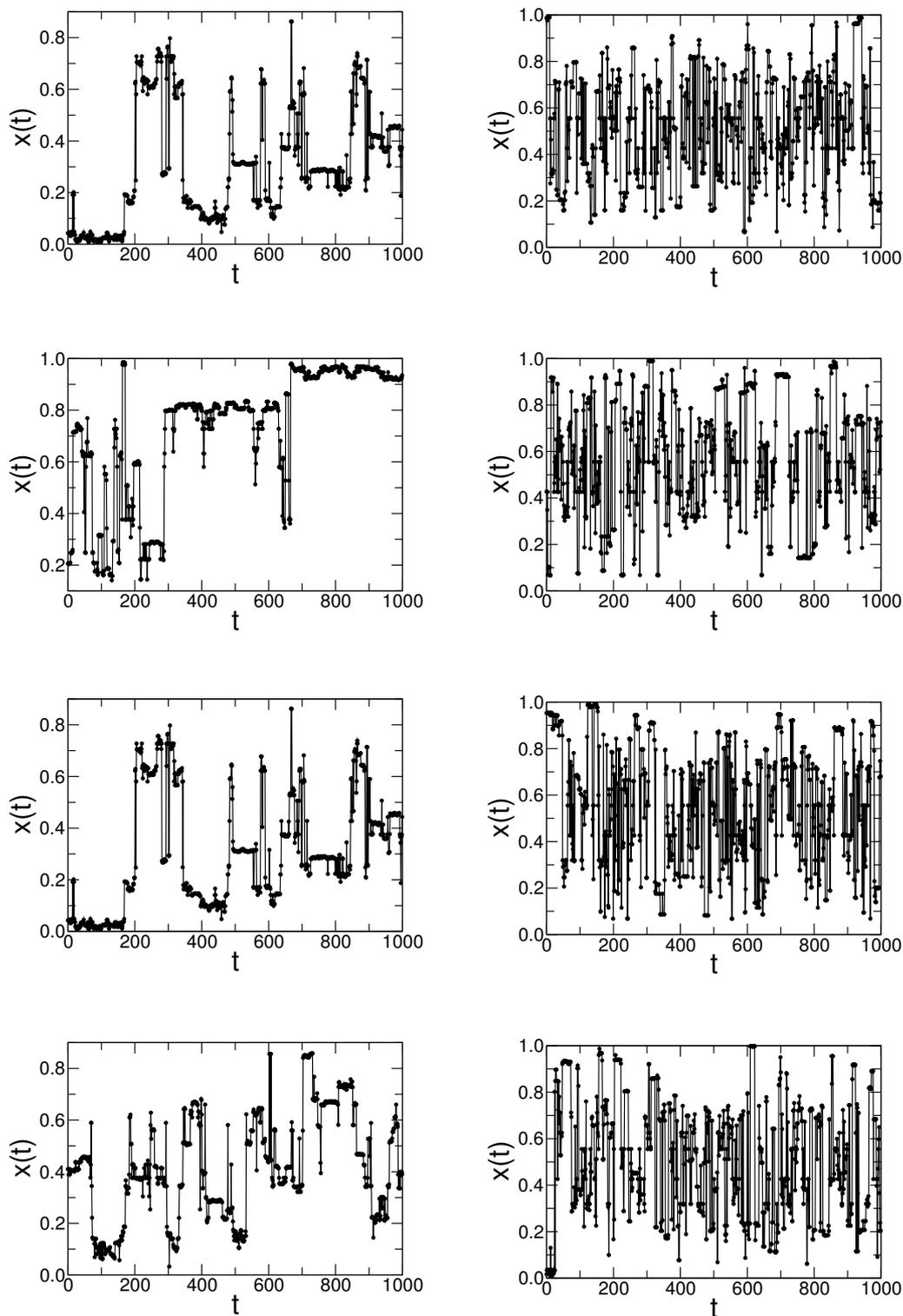


Figura 4.15 - Quatro realizações do mapeamento inverso \mathcal{M}_{QT}^{-1} na rede metabólica *Arabidopsis thaliana* ($Q = 607$ e $T = 100,000$) e na rede de Internet nos Estados Unidos em 1997 ($Q = 1,589$ e $T = 100,000$). Note a similaridade destas séries com as séries apresentadas na Figura 4.13, demonstrando a robustez do mapeamento inverso proposto.

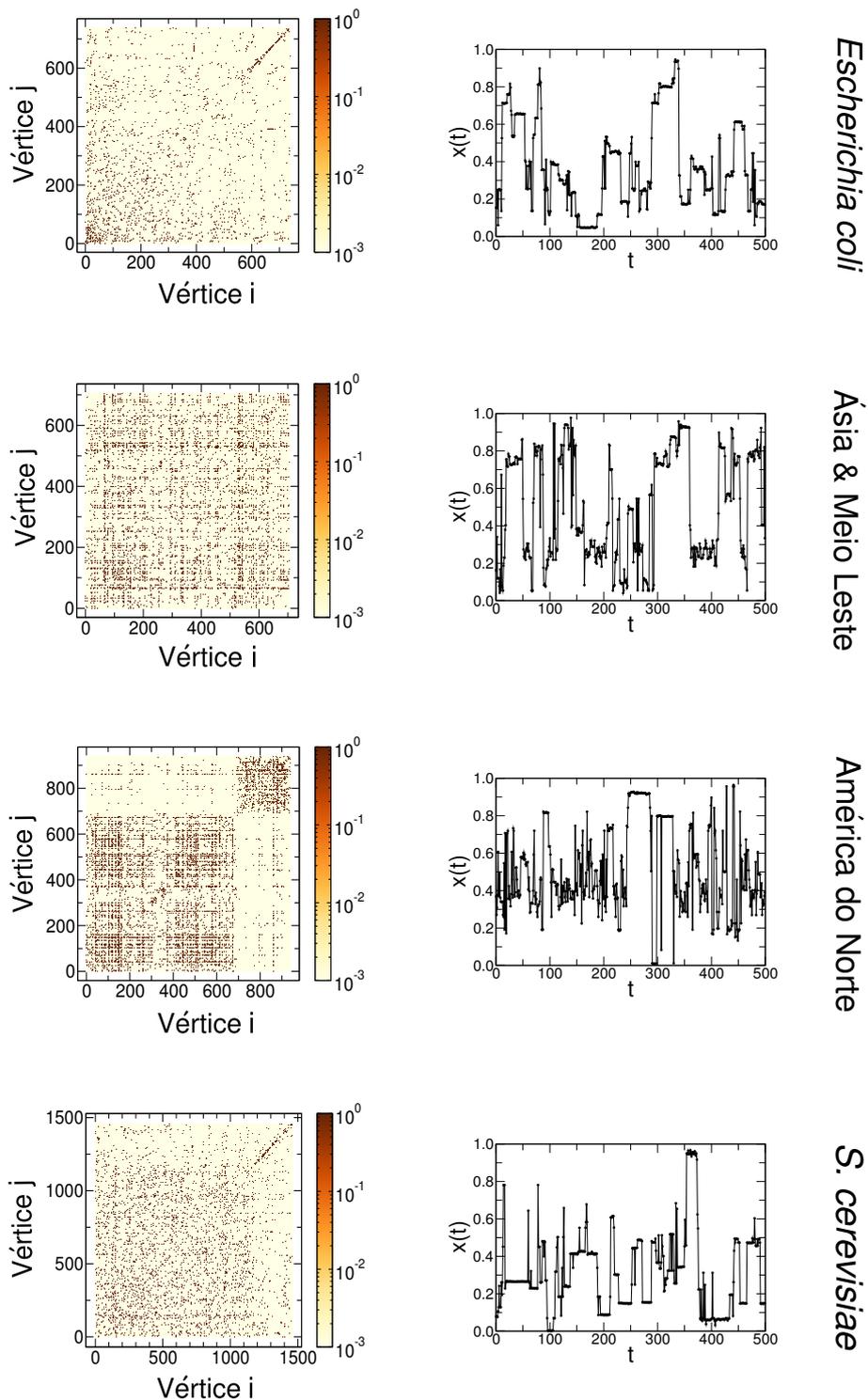


Figura 4.16 - Aplicação do mapeamento inverso em redes complexas reais pertencentes a diferentes classes — a rede metabólica *Escherichia coli*, as redes de transporte aéreo da Ásia & Meio Leste e Norte Americana e a rede de iterações entre proteínas em *Saccharomyces cerevisiae*. Primeiramente, as matrizes de adjacência com pesos associadas são reordenadas e a associação entre vértices e quantis é estabelecida. As séries temporais resultantes apresentam diferenças em dinâmica.

Tipo de rede	Rede	Vértices	Arestas	M_o	$\langle M_o \rangle_{\mathcal{D}}$	H_u	β
Rede Metabólica	<i>Arabidopsis thaliana</i>	607	792	0.825	0.728(0.003)	0.8579(0.0006)	-1.4541(0.0032)
	<i>Escherichia coli</i>	739	1,009	0.0018	0.711(0.003)	0.8138(0.0016)	-1.4031(0.0022)
Rede de Transporte Aéreo	Ásia & Meio Leste	706	2,572	0.642	0.325(0.002)	0.7134(0.0025)	-1.3617(0.0014)
	América do Norte	940	3,446	0.0039	0.3111(0.0005)	0.8089(0.0016)	-1.2329(0.0031)
Rede de Interações entre Proteínas	<i>S. cerevisiae</i>	1,458	1,948	0.820	0.707(0.002)	0.7480(0.0037)	-1.3239(0.0018)
Internet	Internet 1997	1,589	5,627	0.6030	0.5347(0.0010)	0.7045(0.0014)	-1.1865(0.0034)

Tabela 4.1 - Características de diversas redes reais estudadas e das séries temporais correspondentes, obtidas através da aplicação do mapeamento inverso \mathcal{M}_{QT}^{-1} .

As figuras deste capítulo foram geradas com o software PyGrace (<http://pygrace.sourceforge.net>) utilizando-se esquemas de cores da ferramenta ColorBrewer (<http://colorbrewer.org>). Além disso, as redes complexas aqui apresentadas foram geradas pelo software Pajek (<http://vlado.fmf.unilj.si/pub/networks/pajek>). A critério de conhecimento, a disposição dos vértices de uma dada rede no plano (x, y) foi determinada, de forma que, as arestas associadas possuíssem aproximadamente o mesmo tamanho e que houvesse um número mínimo de intersecções entre as mesmas. Mais especificamente, foi utilizado um algoritmo de força direcionada que considera as arestas como objetos elásticos e os vértices como partículas eletricamente carregadas. Desta forma, a rede complexa é tratada como um sistema físico e as forças aplicadas nos vértices fazem com que os mesmos se aproximem ou se afastem ao longo do tempo. Este procedimento é repetido iterativamente até que o sistema atinja um estado de equilíbrio, ou seja, até que as posições dos vértices não se alterem no plano (x, y) de uma iteração para outra (BATAGELJ; MRVAR, 1998).

5 CONCLUSÕES

Os resultados apresentados nesta tese constroem uma ponte conectando a análise de séries temporais e a pesquisa em redes complexas. Neste sentido, uma rede complexa pode ser analisada explorando um conjunto extenso de propriedades estatísticas da série temporal associada. Por exemplo, subgrafos em uma rede serão mapeados em periodicidades em uma série temporal, nas quais podem ser facilmente caracterizadas através do cálculo do espectro de potência associado. Por outro lado, regimes dinâmicos distintos em séries temporais podem ser analisados explorando-se um conjunto extenso de propriedades topológicas das redes associadas (conforme Fig. 4.4, por exemplo).

É importante mencionar que o mapeamento proposto remete algumas técnicas de análise clássica. Por exemplo, o mapeamento direto de uma série temporal em uma rede complexa pode ser comparado com a dinâmica simbólica, onde um sistema contínuo é discretizado em uma sequência de símbolos representando um estado de um sistema dinâmico (LIND; MARCUS, 1995). No mapeamento proposto nesta tese os vértices desempenham o papel de símbolos e uma série simbólica é então obtida através de um caminho pela rede complexa. O mapeamento proposto também fornece uma aproximação única para a compressão de uma série temporal. Desde que a grande parte das séries temporais reais, incluindo séries temporais financeiras e climáticas possuem milhões de pontos, este mapeamento produz naturalmente um mecanismo de compressão de T pontos de uma série temporal em uma lista de no máximo Q^2 elementos da matriz de transição de Markov W .

Finalmente, o mapeamento proposto pode ser modificado a fim de incluir correlações de ordens elevadas. Por exemplo, correlações de segunda ordem podem ser capturadas através do cálculo da densidade de probabilidade markoviana $p(x(t_0 + 1)|x(t_0), x(t_0 - 1))$, o que resultará em redes complexas com hiperconexões (com peso e direcionadas) conectando os vértices associados com os quantis de $x(t_0)$ e $x(t_0 - 1)$ com o vértice associado com o quantil de $x(t_0)$.

O estudo dos tópicos acima (dinâmica simbólica, compressão de dados e captura de correlações de ordem superior), bem como uma investigação mais aprofundada das propriedades das séries temporais e redes reais estudadas no Capítulo 4 constituem sugestões naturais de trabalhos futuros.

6 APÊNDICE A - O algoritmo de Floyd-Warshall

Sejam $\mathcal{N} = \{n_1, n_2, \dots, n_N\}$ os vértices de uma rede complexa direcionada $g = (\mathcal{N}, \mathcal{L})$, e considere um subconjunto $\{n_1, n_2, \dots, n_k\}$ de vértices para algum k . Para quaisquer pares de vértices $n_i, n_j \in \mathcal{N}$, considere todos os caminhos desde n_i até n_j cujos vértices intermediários são todos traçados a partir de $\{n_1, n_2, \dots, n_k\}$, e seja P um caminho de peso mínimo dentre eles. O algoritmo de Floyd-Warshall explora um relacionamento entre o caminho P e caminhos mais curtos desde n_i até n_j com todos os vértices intermediários no conjunto $\{n_1, n_2, \dots, n_{k-1}\}$. O relacionamento depende do fato de n_k ser ou não um vértice intermediário do caminho P .

- Se n_k não é um vértice intermediário do caminho P , então todos os vértices intermediários do caminho P estão no conjunto $\{n_1, n_2, \dots, n_{k-1}\}$. Deste modo, um caminho mais curto desde o vértice n_i até o vértice n_j com todos os vértices intermediários no conjunto $\{n_1, n_2, \dots, n_{k-1}\}$ também é um caminho mais curto desde n_i até n_j com todos os vértices intermediários no conjunto $\{n_1, n_2, \dots, n_k\}$.
- Se n_k é um vértice intermediário do caminho P , então desmembra-se P em n_i até n_k (por P_1) e de n_k até n_j (por P_2). Como o vértice n_k não é um vértice intermediário do caminho P_1 , então P_1 é um caminho mais curto desde n_i até n_k com todos os vértices intermediários no conjunto $\{n_1, n_2, \dots, n_{k-1}\}$. De modo análogo, P_2 é um caminho mais curto desde o vértice n_k até o vértice n_j com todos os vértices intermediários no conjunto $\{n_1, n_2, \dots, n_{k-1}\}$.

Com base nas observações anteriores, define-se uma formulação recursiva de valores de caminhos mais curtos. Seja $d_{ij}^{(k)}$ o peso de um caminho mais curto desde o vértice n_i ao vértice n_j para o qual todos os vértices intermediários estão no conjunto $\{n_1, n_2, \dots, n_k\}$. Quando $k = 0$, um caminho desde o vértice n_i até o vértice n_j sem vértices intermediários com numeração mais alta que 0 não tem absolutamente nenhum vértice intermediário. Tal caminho tem no máximo uma aresta (ou ligação), e então $d_{ij}^{(0)} = w_{ij}$. Uma definição recursiva que segue a discussão anterior é dada por:

$$d_{ij}^{(k)} = \begin{cases} w_{ij} & \text{se } k = 0 \\ \min(d_{ij}^{(k-1)}, d_{ik}^{(k-1)} + d_{kj}^{(k-1)}) & \text{se } k \geq 1 \end{cases}$$

onde os w_{ij} são elementos da matriz W , que representa os pesos das arestas da rede, ou seja:

$$w_{ij} = \begin{cases} 0 & \text{se } i = j \\ \text{o peso da aresta } (n_i, n_j) & \text{se } i \neq j \text{ e } (n_i, n_j) \in K \\ \infty & \text{se } i \neq j \text{ e } (n_i, n_j) \notin K \end{cases}$$

Considerando-se que, para qualquer caminho, todos os vértices intermediários estão no conjunto $\{n_1, n_2, \dots, n_N\}$, a matriz $D^{(N)} = (d_{ij}^{(N)})$ fornece a resposta final: $d_{ij}^{(N)} = \delta(i, j)$ para todo $n_i, n_j \in \mathcal{N}$ (LEISERSON et al., 2002).

O algoritmo de Floyd-Warshall possui diversas características, dentre elas:

- Utiliza uma representação por matriz de adjacências, que é uma estrutura de dados bastante simples;
- É muito eficiente em termos de espaço requerido para armazenamento. A implementação pode utilizar a própria matriz de adjacências para atualizar distâncias;
- É um algoritmo de complexidade cúbica, ou seja, $\mathcal{O}(N^3)$.

7 APÊNDICE B - Reordenamento da matriz de adjacência com pesos

Neste trabalho, a correspondência entre um vértice de uma rede complexa a um quantil na série temporal é feita através do reordenamento da matriz de adjacência com pesos W , de forma que a mesma possua seus maiores elementos w_{ij} próximos da diagonal (SALES-PARDO et al., 2007). Para encontrar um reordenamento próximo ao ótimo, utiliza-se a técnica do recozimento simulado com uma função custo que atribui um peso para cada elemento de W com base na sua distância à diagonal:

$$\mathcal{C} = \frac{1}{N} \sum_{i,j=1}^N W_{ij} |i - j|, \quad (7.1)$$

onde N é a ordem da matriz W .

Para cada interação da busca no recozimento simulado, é proposto $\mathcal{O}(N^2)$ movimentos. Cada tentativa de movimento é realizada ou não utilizando-se o algoritmo Metrópolis. Em seguida, calcula-se o valor da função custo para a nova ordem de \mathcal{C} , definida como \mathcal{C}' , e esta nova ordem é aceita com probabilidade $p = \exp[(\mathcal{C} - \mathcal{C}')/T]$. Inicialmente, a matriz \mathcal{C} é reordenada aleatoriamente a uma temperatura inicial T . Em cada iteração, o valor de T decresce por um fator no domínio $[0.95, 0.999]$ e o processo de reordenamento é finalizado quando os valores de \mathcal{C} se mantêm inalterados após 20 iterações. Informações detalhadas sobre o processo de reordenamento da matriz \mathcal{C} podem ser encontradas em Sales-Pardo et al. (2007).

REFERÊNCIAS BIBLIOGRÁFICAS

- ALBERT, R.; BARABÁSI, A. L. Statistical mechanics of complex networks. **Reviews of Modern Physics**, v. 74, 2002. 2, 10, 15
- ALBERT, R.; JEONG, H.; BARABÁSI, A. L. Diameter of the world wide web. **Nature**, v. 401, 1999. 2
- _____. Evolution of the social network of scientific collaborations. **Physica A**, v. 311, 2002. 1
- ALLIGOOD, K. T.; SAUER, T. D.; YORK, J. A. **Chaos: an introduction to dynamical systems**. New York: Springer-Verlag, 1997. 23, 24
- AMARAL, L.; MEYER, M. Environmental changes, coextinction, and patterns in the fossil record. **Physical Review Letters**, v. 82, 1999. 2
- AMARAL, L. A. N.; OTTINO, J. Complex networks: Augmenting the framework for the study of complex systems. **The European Physical Journal B**, v. 38, 2004. 5
- AO, S. **Applied time series analysis and innovative computing**. New York: Springer Publishing Company, Incorporated, 2010. 3, 29
- ARGYRIS, J.; FAUST, G.; HAASE, M. **An exploration of chaos**. New York: Elsevier, 1994. 27
- BAKER, G. L.; GOLLUB, J. P. **Chaotic dynamics: an introduction**. New York: Cambridge University Press, 1996. 26
- BARABÁSI, A.-L. **Linked: how everything is connected to everything else and what it means for business, science, and everyday life**. New York: Plume, 2002. 1, 14
- BARABÁSI, A. L.; ALBERT, R. Emergence of scaling in random networks. **Science**, v. 289, 1999. xxv, 19
- BATAGELJ, V.; MRVAR, A. **PAJEK**: program for large network analysis. 1998. Disponível em: <<http://vlado.fmf.uni-lj.si/pub/networks/pajek/>>. 62

- BOCCALETTI, S.; LATORA, V.; MORENO, Y.; CHAVEZ, M.; HWANG, D. U. Complex networks:: Structure and dynamics. **Physics Reports**, v. 424, 2006. 1, 6, 8, 11
- BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. **Time series analysis, forecasting and control**. New York: John Wiley & Sons, 2008. 3
- CAMACHO, J.; STOUFFER, D. B.; AMARAL, L. A. N. Quantitative analysis of the local structure of food webs. **Journal of Theoretical Biology**, v. 246, 2007. 2
- CAMPANHARO, A. S. L. O.; RAMOS, F. M.; MACAU, E. E. N.; ROSA, R. R.; BOLZAN, M. J. A.; SÁ, L. D. A. Searching chaos and coherent structures in the atmospheric turbulence above the amazon forest. **Phil. Trans. R. Soc. A**, v. 366, 2008. 3
- COOLEY, J.; TUKEY, J. An algorithm for the machine calculation of complex fourier series. **Mathematics of Computation**, v. 19, n. 90, p. 297–301, 1965. 26
- COSTA, L. F.; RODRIGUES, F. A.; TRAVIESO, G.; VILLAS, P. R. Characterization of complex networks: A survey of measurements. **Advances in Physics**, v. 56, 2007. 3, 6, 8, 12, 14
- ERDŐS, P.; RÉNYI, A. On random graphs. **Publicationes Mathematicae**, v. 6, 1959. xxv, 14
- FAGIOLO, G. Clustering in complex directed networks. **Physical Review E**, v. 76, 2007. 12
- FALOUTSOS, M.; FALOUTSOS, P.; FALOUTSOS, C. On power-law relationships of the internet topology. **Computer Communication Review**, v. 29, 1999. 2
- GAO, Z.; JIN, N. Complex network from time series based on phase space reconstruction. **Chaos**, v. 19, 2009. 3, 33, 35
- GLEISER, P. M.; DANON, L. Community structure in jazz. **Advances in Complex Systems**, v. 6, n. 4, 2003. 1
- GOLDBERGER, A. L.; AMARAL, L. A. N.; GLASS, L.; HAUSDORFF, J. M.; IVANOV, P. C.; MARK, R. G.; MIETUS, J. E.; MOODY, G. B.; PENG, C.-K.; STANLEY, H. E. Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. **Circulation**, v. 101, 2000. 51

- GUIMERÀ, R.; SALES-PARDO, M.; AMARAL, L. Modularity from fluctuations in random graphs and complex networks. **Physical Review E**, v. 70, 2004. 13
- GUIMERÀ, R.; SALES-PARDO, M.; AMARAL, L. Classes of complex networks defined by role-to-role connectivity profiles. **Nature Physics**, v. 3, p. 63–69, 2007. 55
- GUIMERÀ, R.; UZZI, B.; SPIRO, J.; AMARAL, L. A. N. Team assembly mechanisms determine collaboration network structure and team performance. **Science**, v. 308, 2005. 1
- HARAGUCHI, Y.; SHIMADA, Y.; IKEGUCHI, T.; AIHARA, K. Transformation from complex networks to time series using classical multidimensional scaling. In: **Proceedings...** Berlin: Springer-Verlag, 2009. 37
- HUBERMAN, B. A. **The laws of the web: patterns in the ecology of information**. Cambridge, USA: MIT Press, 2001. 2
- HURST, H. Long-term storage capacity of reservoirs. **Trans. Amer. Soc. Civil Eng**, v. 110, 1951. 29
- KANTZ, H.; SCHREIBER, T. **Nonlinear time series analysis**. Cambridge: Cambridge University Press, 2003. 3, 24
- KIRKPATRICK, S.; GELATT, C. D.; VECCHI, M. P. Optimization by simulated annealing. **Science**, v. 220, p. 671–680, 1983. 37
- KORNER, T. W. **Fourier analysis**. Cambridge: Cambridge University Press, 1988. 3
- KRAUSE, A.; FRANK, K. A.; MASON, D. M.; ULANOWICZ, R. E.; TAYLOR, W. M. Compartments exposed in food-web structure. **Nature**, v. 426, 2003. 2
- 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**, v. 105, 2008. xxv, 3, 32, 33
- LAI, C.; CHUNG, P.; TSENG, V. S. A novel two-level clustering method for time series data analysis. **Expert Systems with Applications**, v. 37, 2010. 3, 29
- LATORA, V.; MARCHIORI, M. Efficient behavior of small-world networks. **Physical Review E**, v. 19, 2001. 13

- LEISERSON, C. E.; CLIFFORD, S.; RIVEST, R. L.; CORMEN, T. H.
Algoritmos: teoria e prática. Rio de Janeiro: Elsevier Editora, 2002. 66
- LILJEROS, F.; EDLING, C. R.; AMARAL, L. A. N.; STANLEY, H. E.; ABERG, Y.
 The web of human sexual contacts. **Nature**, v. 411, 2001. 1
- LIND, D.; MARCUS, B. **An introduction to symbolic dynamics and coding.** Cambridge: Cambridge University Press, 1995. 63
- LORENZ, E. N. Deterministic nonperiodic flow. **Journal of Atmospheric Sciences**, v. 20, 1963. 24, 44
- LUQUE, B.; LACASA, L.; BALLESTEROS, F.; LUQUE, J. Horizontal visibility graphs: exact results for random time series. **Physical Review E**, v. 80, 2009. 3, 33
- MALAMUD, B. D.; TURCOTTE, D. Self-affine time series: measures of weak and strong persistence. **Journal of Statistical Planning and Inference**, v. 80, 1999. 46
- MARWAN, N.; DONGES, J. F.; ZOU, Y.; DONNER, R. V.; KURTHS, J.
 Complex network approach for recurrence analysis of time series. **Physics Letters A**, v. 46, 2009. 3, 33, 35
- MASLOV, S.; SNEPPEN, K. Specificity and stability in topology of protein networks. **Science**, v. 296, n. 5569, 2002. 2
- MATIA, K.; AMARAL, L. A. N.; LUWEL, M.; MOED, H. F.; STANLEY, H. E.
 Scaling phenomena in the growth dynamics of scientific output. **Journal of the American Society for Information Science and Technology**, v. 56, 2005. 1
- MENDES, J. F. F. Física de redes complexas. **Gazeta de Física**, Coimbra, 2005. 7
- MILGRAN, S. The small world problem. **Psychology Today**, v. 1, 1967. 15
- NETGEO - The Internet Geographic Database.
[Http://www.caida.org/tools/utilities/netgeo](http://www.caida.org/tools/utilities/netgeo). 55
- NEWMAN, M. E. J. Structure and function of complex networks. **SIAM Review**, v. 45, 2003. 3, 10

_____. Modularity and community structure in networks. **Proceedings of the National Academy of Sciences**, v. 103, n. 23, 2006. 14

ONODY, R. N.; CASTRO, P. A. Complex network study of brazilian soccer players. **Physical Review E**, v. 70, 2004. 1

PAPOULIS, A. **The Fourier integral and its applications**. New York: McGraw Hill, 1962. 25

PERCIVAL, D. B.; WALDEN, A. T. **Wavelet methods for time series analysis**. Cambridge: Cambridge University Press, 2000. 3

PHYSIONET: the research resource for complex physiologic signals.
[Http://www.physionet.org](http://www.physionet.org). 51

REDNER, S. How popular is your paper: An empirical study of the citation distribution. **The European Physical Journal B - Condensed Matter and Complex Systems**, v. 4, n. 2, 1998. 1

RODRIGUES, F. A. **Caracterização, classificação e análise de redes complexas**. Tese (Tese de Doutorado) — Universidade de São Paulo, São Carlos, 2007. 9, 17, 20, 22

ROSSLER, O. E. An equation for continuous chaos. **Physics Letters**, v. 57A, 1976. 44

RUGGIERO, C. A.; BRUNO, O. M.; TRAVIESO, G.; COSTA, L. F. On the efficiency of data representation on the modeling and characterization of complex networks. **Physica A**, v. 390, n. 11, 2011. 10

SALES-PARDO, M.; GUIMERÀ, R.; MOREIRA, A. A.; AMARAL, L. A. N. Extracting the hierarchical organization of complex systems. **Proceedings of the National Academy of Sciences**, v. 104, p. 15224–15229, 2007. 13, 37, 38, 55, 67

STOUFFER, D. B.; CAMACHO, J.; JIANG, W.; N., A. L. A. Evidence for the existence of a robust pattern of prey selection in food webs. **Proceedings of the Royal Society B**, v. 274, 2007. 2

STOUFFER, D. B.; CAMACHO, J.; N., A. L. A. A robust measure of food web intervality. **Proceedings of the National Academy of Sciences**, v. 103, 2006. 2

- STROGATZ, S. H. **Nonlinear dynamics and chaos**: with applications to physics, biology, chemistry, and engineering. Boulder: Westview Press, 1994. 3
- STROZZI, F.; ZALDÍVAR, J.-M.; POLJANSEK, K.; BONO, F.; GUTIÉRREZ, E. From complex networks to time series and vice versa : Application to metabolic networks. **Communities**, EUR 23947, 2009. 37
- VERPLANCKE, T.; LOOY, S. V.; STEURBAUT, K.; BENOIT, D.; TURCK, F. D.; MOOR, G. D.; DECRUYENAERE, J. A novel time series analysis approach for prediction of dialysis in critically ill patients using echo-state networks. **BMC Medical Informatics and Decision Making**, v. 10, 2010. 3, 29
- WASSERMAN, S.; FAUST, K. **Social network analysis**: methods and applications. New York: Cambridge University Press, 1994. 11
- WATTS, D. J.; STROGATZ, S. H. Collective dynamics of “small world” networks. **Letters to Nature**, v. 393, 1998. xxv, 11, 13, 15, 19, 39
- XU., X.; ZHANG, J.; SMALL, M. Superfamily phenomena and motifs of networks induced from time series. **Proc Natl Acad Sci U S A**, v. 105, 2008. 3, 33, 35
- YANG, Y.; YANG, H. J. Complex network-based time series analysis. **Physica A**, v. 387, 2008. 3, 33, 35
- ZHANG, J.; LUO, X.; SMALL, M. Detecting chaos in pseudoperiodic time series without embedding. **Physical Review E**, v. 73, 2006. 3, 29
- ZHANG, J.; SMALL, M. Complex network from pseudoperiodic time series: Topology versus dynamics. **Physical Review Letters**, v. 96, 2006. xxv, 3, 29, 31, 33

PUBLICAÇÕES TÉCNICO-CIENTÍFICAS EDITADAS PELO INPE

Teses e Dissertações (TDI)

Teses e Dissertações apresentadas nos Cursos de Pós-Graduação do INPE.

Manuais Técnicos (MAN)

São publicações de caráter técnico que incluem normas, procedimentos, instruções e orientações.

Notas Técnico-Científicas (NTC)

Incluem resultados preliminares de pesquisa, descrição de equipamentos, descrição e ou documentação de programas de computador, descrição de sistemas e experimentos, apresentação de testes, dados, atlas, e documentação de projetos de engenharia.

Relatórios de Pesquisa (RPQ)

Reportam resultados ou progressos de pesquisas tanto de natureza técnica quanto científica, cujo nível seja compatível com o de uma publicação em periódico nacional ou internacional.

Propostas e Relatórios de Projetos (PRP)

São propostas de projetos técnico-científicos e relatórios de acompanhamento de projetos, atividades e convênios.

Publicações Didáticas (PUD)

Incluem apostilas, notas de aula e manuais didáticos.

Publicações Seriadas

São os seriados técnico-científicos: boletins, periódicos, anuários e anais de eventos (simpósios e congressos). Constam destas publicações o Internacional Standard Serial Number (ISSN), que é um código único e definitivo para identificação de títulos de seriados.

Programas de Computador (PDC)

São a seqüência de instruções ou códigos, expressos em uma linguagem de programação compilada ou interpretada, a ser executada por um computador para alcançar um determinado objetivo. Aceitam-se tanto programas fonte quanto os executáveis.

Pré-publicações (PRE)

Todos os artigos publicados em periódicos, anais e como capítulos de livros.