#### UNIVERSIDADE DE SÃO PAULO INSTITUTO DE FÍSICA DE SÃO CARLOS

ANDRÉ LUIZ BARBIERI

#### Análise de Robustez em Redes Complexas

São Carlos 2010

#### ANDRÉ LUIZ BARBIERI

#### Análise de Robustez em Redes Complexas

Dissertação apresentada ao Programa de Pósgraduação em Física do Instituto de Física de São Carlos da Universidade de São Paulo para a obtenção do título de Mestre em Ciência.

Área de Concentração: Física Computacional Orientador: Prof. Dr. Francisco Aparecido Rodrigues

Versão Original

São Carlos 2010

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

Ficha catalográfica elaborada pelo Serviço de Biblioteca e Informação IFSC/USP

Barbieri, Andre Luiz. Análise de robustez em redes complexas / Andre Luiz Barbieri; orientador Francisco Aparecido Rodrigues. -- São Carlos, 2011. 110 p.

Dissertação (Mestrado em Ciências - Área de concentração: Física Aplicada) – Instituto de Física de São Carlos da Universidade de São Paulo.

1. Redes complexas. 2. Robustez. 3. Proteínas. 4. Acessibilidade. 5. Rodovias. I. Título.

#### FOLHA DE APROVAÇÃO

André Luiz Barbieri

Dissertação apresentada ao Instituto de Física de São Carlos da Universidade de São Paulo para obtenção do título de Mestre em Ciências. Área de Concentração: Física Aplicada -Opção: Física Computacional

Aprovado(a) em: 14.02.2011

Comissão Julgadora

Prof(a). Dr(a). Francisco Aparecido Rodrigues

Instituição: ICMC/USP

Assinatura france Su

Prof(a). Dr(a). Virgilio Augusto Fernandes Almeida Assinatura\_ Instituição: UFMG

Prof(a). Dr(a). Zhao Liang Instituição: ICMC/USP

Assinatura

hading

À minha esposa, Adriele, com muito amor, carinho, admiração e gratidão pela sua compreensão, presença e incansável apoio ao longo do período de elaboração deste trabalho.

#### **AGRADECIMENTOS**

Ao pesquisador, amigo e orientador Francisco A. Rodrigues, que durante todos estes anos colaborou de forma positiva com a minha formação científica e intelectual, sempre acreditando em mim em todos os momentos.

Ao professor Luciano da F. Costa, por me receber de braços abertos em seu grupo de pesquisa.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela concessão da bolsa de estudos.

Ao Instituto de Física de São Carlos, por me acolher de maneira receptiva e calorosa durante toda a minha formação.

Aos grandes amigos, Paulino, *Tiozão* (Gerson), *Rato* (Matheus) e Renato, que foram como irmãos durante todo o meu percurso na faculdade, aos grandes parceiros, *Cuiabá* (João Bunoro) e *Don Dieguito* (Diego), pelo grande incentivo que me deram neste trabalho e pelos grandes passes no futebol, ao *Porks* (Marcos Felipe) pelas discussões construtivas sobre a ciência e religião, e ao *Presidente* (Fernando) pela auxílio na finalização deste projeto.

À minha Mãe, Ana, e ao meu Pai, Pedro, por tudo que fizeram por mim, pela valiosa educação que sempre me proporcionaram, pelo carinho e amor eternos, e pelos sábios conselhos que sempre me lembrarei.

Às bibliotecárias do SBI-IFSC, Lena, Neusa, Mara, Marilza, Sy, Lu, Cris, Be, Célia, Nata, pela paciência que tiveram comigo, nos vários anos em que trabalhei junto destas lindas mulheres.

E a todos que, de alguma maneira ou outra, contribuíram para a elaboração deste trabalho.

"Se vi mais longe foi por estar sobre os ombros de gigantes." — ISAAC NEWTON (1643 - 1727)

"A coisa mais incompreensível sobre o mundo é que ele é compreensível a todos." — ALBERT EINSTEIN (1879 - 1955)

#### **RESUMO**

BARBIERI, A. L. **Análise de Robustez em Redes Complexas.** 2011. 110 p. Dissertação (Mestrado) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2011.

A teoria das redes complexas é uma área relativamente nova da Ciência, inspirada por dados empíricos tais como os obtidos de interações biológicas e sociais. Esta área apresenta uma natureza altamente interdisciplinar, de modo que tem unido cientistas de diferentes áreas, tais como matemática, física, biologia, ciência computação, sociologia, epidemiologia e muitas outras. Um dos problemas fundamentais nessa área é entender como a organização de redes complexas influencia em processos dinâmicos, como sincronização, propagação de epidemias e falhas e ataques. Nessa dissertação, é apresentada uma análise da relação entre estrutura e robustez de redes complexas através da remoção de vértices. Para a aplicação deste estudo, foram adquiridas bases de dados de interações de proteínas de quatro espécies, Saccharomyces cerevisiae, Caenorhabditis elegans, Drosophila melanogaster e Homo sapiens, como também mapas das malhas de rodovias de sete países, Brasil, Portugal, Polônia, Romênia, Austrália, Índia e África do Sul. Foi estudada a robustez dessas redes através de simulação de falhas e ataques, segundo uma dinâmica de remoção de vértices. Nesse caso, a variação na estrutura das redes devido a essa remoção foi quantificada pelas medidas do tamanho da maior componente conectado, do diâmetro e da média dos comprimentos dos menores caminhos. Ademais, foram utilizadas duas medidas para quantificar a robustez, isto é, a entropia da distribuição das conexões e entropia dinâmica, baseada em cadeias de Markov. Tais medidas foram aplicadas nas redes reais onde se verificou que as espécies mais complexas, como o homem e a mosca, apresentam as redes mais robustas. Com relação aos países, Romênia, Portugal e Brasil apresentam as malhas rodoviárias mais resistentes a perturbações. A correlação entre essas medidas de entropia e as medidas topológicas permitiu identificar que a média do grau dos vizinhos e o coeficiente da lei de potência da distribuição do número de conexões são as medidas que apresentam maior correlação com as medidas de entropia. Tal resultado sugere que a presença de conexões alternativas entre os vizinhos dos vértices removidos favorece a resiliência das redes, pois tendem a minimizar as perturbações causadas pelas remoções. No caso das malhas rodoviárias, foi proposta uma nova medida de acessibilidade e esta se mostrou altamente correlacionada com a entropia dinâmica. Nesse caso, verificou-se que as cidades localizadas no litoral e nas fronteiras dos países são as que menos contribuem para robustez das redes de rodovias. Desse modo, os resultados obtidos sugerem que o planejamento do sistema de transporte de um país deve priorizar o investimento em infra-estrutura rodoviária próximo das cidades com menor acessibilidade, de forma a tornálas mais acessíveis, visando melhorar o transporte de mercadorias e pessoas. Os métodos aqui propostos permitem identificar tais cidades. Ademais, na análise de redes de proteínas, os resultados obtidos podem auxiliar no desenvolvimento de novos modelos de redes, bem como entender os mecanismos evolutivos que priorizam a robustez dos organismos.

Palavras-chave: Redes complexas. Robustez. Proteínas. Acessibilidade. Rodovias.

### ABSTRACT

BARBIERI, A. L. **Analysis of Robustness in Complex Networks.** 2011. 110 p. Dissertation (Master) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2011.

The study of complex networks is a relatively new area of science inspired by the empirical studies of real-world networks, such as social and biological networks. This are has a highly multidisciplinary nature, which has brought together researchers from many areas including mathematics, physics, biology, computer science, sociology, epidemiology, statistics and others. One of the main problems in this area is to know how the network organization is related to dynamic process, such as synchronization, epidemic spreading and topological perturbation due to deletion of nodes and edges. In this dissertation, it is presented a study of the relationship between the structure and resilience of complex networks. This investigation was applied to the protein-protein networks of four species, namely Saccharomyces cerevisiae, Caenorhabditis elegans, Drosophila melanogaster and Homo sapiens, as well as the road networks of seven countries, i.e. Brazil, Portugal, Romania, Australia, India, and South Africa. It was studied the resilience of such networks through simulations of random fails and attacks by node deletion. The topological changes due to this simulation were quantified by measures, including the size of the largest component, the diameter and the average shortest path length. In addition, the network robustness was quantified by the entropy of the degree distribution and the dynamic entropy, related to Markov chains. This analysis in real-world networks revealed that more complex species, such as the H. sapiens and D. melanogaster are the most resilient. In addition, Romania, Brazil and Portugal have the most robust road maps. The correlation analysis between topological and dynamic measures revealed that the average neighborhood degree and the coefficient of scaling in the power law of the degree distribution quantify the proprieties that most contribute for the resilience in protein networks. Moreover, with respect to the road networks, it was introduced a new accessibility measure, which revealed to be correlated to the dynamic entropy. In fact, cities localized in the border of networks are the ones with the smallest contribution for the network resilience. Therefore, the obtained results suggest that the traffic planning should mainly connect cities near the frontiers of countries, in order to improve the resilience and accessibility. In addition, the obtained results with respect to protein networks allow improving network modeling and understanding the biological processes that reinforce the resilience of organisms.

Keywords: Complex networks. Robustness. Proteins. Accessibility. Roads.

### LISTA DE FIGURAS

Figura 2.1 -	Diagrama das pontes da cidade de Königsberg	39
Figura 2.2 -	Representação em (a) de uma rede não dirigida de 6 vértices, e em (b) seu armazenamento computacional na forma de lista de conexões	41
Figura 2.3 -	Representação de uma rede dirigida de 6 vértices	43
Figura 2.4 -	Representação em (a) de uma rede com pesos em suas ligações, e seu armazenamento na forma de lista de conexões em (b)	44
Figura 2.5 -	<ul><li>(a) Rede social de amizades em um clube de karate (1). (b) Histograma</li><li>da distribuição de conectividade normalizada</li></ul>	46
Figura 2.6 -	Exemplificação para os valores de coeficiente de aglomeração calcula- dos para o vértice em vermelho.	47
Figura 2.7 -	Rede geográfica de rodovias do Brasil. Os pontos representam as cidades.	48
Figura 2.8 -	(a) Exemplo de uma rede aleatória de Erdös e Rényi. (b) Distribuição da conectividade para uma rede com N=1000, usando uma probabilidade $p = 0, 2$ . Os pontos no gráfico são a média de 100 simulações	50
Figura 2.9 -	Exemplificação da variação da probabilidade $p$ no modelo de Watts-	

Strogatz com conectividade média  $\langle k \rangle = 4....51$ 

Figura 2.10 -	(a) Gráfico do coeficiente de aglomeração $cc$ e do caminho médio $\ell$ para	
	os diferentes valores de probabilidade $p$ na geração de redes do modelo	
	WS (2). (b) Distribuição da conectividade para uma rede com N=1000,	
	$\langle k \rangle = 200$ , usando uma probabilidade $p = 0,01$ . Os pontos no gráfico	
	são a média de 100 simulações	52
Figura 2.11 -	(a) Exemplo de uma rede livre de escala gerada pelo modelo $SF$ . (b)	
	Gráfico em escala logarítmica da distribuição de conectividade (em azul)	
	e da distribuição cumulativa (em vermelho) para uma rede com N=1000.	
	Os pontos no gráfico são a média de 100 simulações	54
Figura 2.12 -	Exemplos de redes complexas: (a) uma amostra da <i>World Wide Web</i> , (b)	
	a Internet, (c) a rede social dos personagens do livro "Os miseráveis",	
	de Victor Hugo, (d) a rede de interações de proteínas da levedura Sac-	
	charomyces cerevisiae	56
Figura 3.1 -	Figura adaptada do artigo (3). Média do grau dos vizinhos mais próximos	
	de vértices com grau k, para a rede da Internet para os anos de 1997 a	
	1999	59
Figura 3.2 -	Figura adaptada do artigo (4). Medida da varição do caminho médio	
	entre os vértices $\ell$ em relação a remoção aleatória dos nós. Os valores	
	da entropia da distribuição de conectividade também são indicados	63
Figura 3.3 -	Simulação de falhas e ataques na Internet: (a) diâmetro da rede, (b)	

- Figura 4.2 Distribuição acumulada do grau das redes de interação de proteína-proteína da *S. cerevisiae* (Sce-CCSB e Sce-Union), *C. elegans* (Cel-Wi e Cel-BG), *D. melanogaster* (Dme-BG e Dme-Fl) e *H. sapiens* (Hsa-BG e Hsa-CCSB).
  75

- Figura 4.6 Distribuição do grau médio dos vizinhos para os hubs removidos da *S. cerevisiae* (Sce-CCSB e Sce-Union), *C. elegans* (Cel-Wi and Cel-BG), *D. melanogaster* (Dme-BG e Dme-Fl) e *H. sapiens* (Hsa-BG e Hsa-CCSB). A mediana de cada distribuição, *m*, também é indicada nos respectivos gráficos.
- Figura 5.2 (a) Mapa de interligação de cruzamentos do Brasil. (b) Cruzamentos rotulados por cores, através do algoritmo de propagação de rótulos. . . . 85

Figura 5.4 -	Entropia dinâmica das redes de rodovias estudadas: (a) Austrália, (b)	
	Brasil, (c) Índia, (d) Romênia, (e) Portugal, (f) Polônia e (g) África do	
	Sul. As cores indicam os valores da entropia dinâmica de cada vértice	
	nas redes.	88
Figura 5.5 -	Ilustração do conceito de acessibilidade. Os valores da acessibilidade de	
	cada vértice estão indicados na figura.	90
Figura 5.6 -	Grau de intermediação para as redes de rodovias estudadas: (a) Austrália,	
	(b) Brasil, (c) Índia, (d) Romênia, (e) Portugal, (f) Polônia e (g) África	
	do Sul. As cores indicam os valores da entropia dinâmica de cada vértice	
	nas redes.	92
Figura 5.7 -	Classificação dos vértices de acordo com as medidas de robustez, aces-	
	sibilidade e centralidade para os países estudados: (a) Austrália, (b) Bra-	
	sil, (c) Índia, (d) Romênia, (e) Portugal, (f) Polônia e (g) África do Sul.	
	As cores indicam os valores da entropia dinâmica de cada vértice nas	
	redes	94

# LISTA DE TABELAS

Tabela 2.1 -	Exemplos de redes complexas reais que seguem uma distribuição livre	
	de escala.	56
Tabela 3.1 -	Assortatividade em redes reais	58
Tabela 4.1 -	Conjunto de medidas obtidas para a redes das quatros espécies consi-	
	deradas, S. cerevisiae (Sce-CCSB e Sce-Union), C. elegans (Cel-Wi e	
	Cel-BG), D. melanogaster (Dme-BG e Dme-Fl) e H. sapiens (Hsa-BG	
	e Hsa-CCSB).	76
Tabela 5.1 -	Propriedades topológicas das redes utilizadas: número de vértices (N),	
	número de conexões ( <i>M</i> ), coeficiente de aglomeração médio ( $\langle cc \rangle$ ), média	
	dos menores caminhos ( $\ell$ ), diâmetro ( $d$ ), média da conectividade média	
	dos vizinhos de cada vértice ( <i>r</i> )	86
Tabela 5.2 -	Valores da entropia dinâmica $(H_d)$ , entropia de conectividade $(H)$ , aces-	
	sibilidade (Acc) e betweenness centrality obtidas para as redes de rodo-	
	vias estudadas	89
Tabela 5.3 -	Valores da correlação entre a entropia dinâmica local $(H_d(i))$ e a acessi-	
	bilidade ( $Acc(i)$ )	90
Tabela 5.4 -	Valores da correlação entre o grau de intermediação $(B_u)$ e a acessibili-	
	dade $(Acc(i))$	91

# LISTA DE ABREVIATURAS

BG	Biogrid database
CCSB	Center for Cancer Systems Biology
Cel	Caenorhabditis elegans
Dme	Drosophila melanogaster
Droid	Drosophila interaction database
ER	Modelo de rede aleatória de Erdõs e Rényi
FI	Finley database
GIS	Sistemas de Informação Geográfica
Hsa	Homo sapiens
HTS	High throughput screening
RNA	Ribonucleic acid
Sce	Saccharomyces cerevisiae
SF	Modelo de rede scale-free de Barabási e Albert
Union	Base de dados de Ito e Uetz
Wi	Worm Interactome database
WS	Modelo de rede small-word de Watts e Strogatz

# LISTA DE SÍMBOLOS

R	Rede Complexa
Ŷ	Conjunto de vértices de uma rede
Ν	Número total de vértices
E	Conjunto de conexões de uma rede
М	Número total de conexões
A	Matriz de adjacência
$a_{ij}$	Elemento da matriz de adjacência
W	Pesos das conexões
Ŵ	matriz de pesos
k <sub>i</sub>	Conectividade do vértice
$\langle k  angle$	Conectividade média
P(k)	Distribuição de conectividade
cci	Coeficiente de aglomeração do vértice
$\langle cc \rangle$	Coeficiente de aglomeração médio
e <sub>i</sub>	Número de ligações entre os vértices da vizinhança
D	Matriz de distâncias
$d_{ij}$	Comprimento do menor caminho

$d_{max}$	Diâmetro da rede
l	Menor caminho médio
γ	Coeficiente da lei de potência
P(k' k)	Probabilidade condicional
$\langle k_{nn} \rangle$	Média do grau dos vizinhos mais próximos de vértices com grau k
S	Medida de assortatividade pelo coeficiente de correlação de Pearson
r(i)	Conectividade média dos vizinhos mais próximos do vértice
r	Média amostral de $r(i)$ entre os vértices da rede
$B_u$	Grau de intermediação
$\sigma(i,u,j)$	Número de menores caminhos entre os vértices $i$ e $j$
В	Média do grau de intermediação
Н	Entropia da distribuição de conectividade $P(k)$
$H_d$	Entropia dinâmica para um processo de Markov
Р	Matriz de probabilidade de transição
ρ	Medida de correlação
$Acc_h(i)$	Acessibilidade do vértice para uma distância h
Acc(i)	Medida de acessibilidade do vértice

# **SUMÁRIO**

1	Intr	rodução 3					
	1.1	Objetivos	. 33				
	1.2	Contribuições	. 34				
	1.3	Descrição dos capítulos	. 35				
2	Con	iceitos básicos	37				
	2.1	O que é uma rede	. 37				
	2.2	Introdução histórica	. 38				
	2.3	Como representar uma rede	. 40				
		2.3.1 Redes não dirigidas	. 41				
		2.3.2 Redes dirigidas	. 42				
		2.3.3 Redes com pesos	. 43				
	2.4	Medidas básicas para caracterização de redes	. 44				
		2.4.1 Conectividade	. 45				
		2.4.2 Distribuição de conectividade	. 45				
		2.4.3 Coeficiente de aglomeração	. 46				
		2.4.4 Distâncias	. 47				
	2.5	Modelos básicos	. 49				

		2.5.1	Redes aleatórias	49		
		2.5.2	Modelo de Watts-Strogatz	51		
		2.5.3	Redes livres de escala	52		
	2.6	Exemp	plos de redes reais	55		
3	Cara	acteriza	ıção de redes complexas	57		
	3.1	Caract	erização topológica de redes	57		
		3.1.1	Assortatividade	58		
		3.1.2	Conectividade média dos vizinhos	60		
		3.1.3	Grau de intermediação	60		
		3.1.4	Entropia da distribuição das conexões	61		
	3.2	Caract	erização da robustez de redes complexas	62		
		3.2.1	Falhas e ataques	64		
		3.2.2	Medida para caracterização da robustez	66		
4	Rob	ustez do	e redes de interação proteína-proteína	69		
	4.1	Redes	biológicas	69		
	4.2	2 Bases de dados				
	4.3	3 Caracterização de redes de interação de proteínas				
		4.3.1	Medidas topológicas	73		
		4.3.2	Medidas relacionadas a resiliência	74		
	4.4	Result	ados obtidos	74		
		4.4.1	Falhas aleatórias	76		

Re	Referências			99
6	Con	clusões	e trabalhos futuros	95
	5.2	Result	ados	85
	5.1	Aquisi	ção de dados	83
5	Rob	ustez, a	cessibilidade e centralidade de malhas rodoviárias	83
		4.4.3	Relação entre a estrutura da rede e a resiliência	78
		4.4.2	Ataque intencional	78

# 1 INTRODUÇÃO

Ao longo do seu desenvolvimento, a pesquisa científica se especializou adotando uma abordagem reducionista, onde sistemas são divididos em partes, que são analisadas individualmente, de modo a entender seu comportamento global. No entanto, apesar do grande avanço obtido, tal abordagem se revelou incapaz de explicar o comportamento da maioria dos sistemas ditos complexos, que são formados de componentes individuais que se influenciam mutuamente (6, 7). Essa limitação do tratamento reducionista pode ser vista, por exemplo, na Internet, onde a dinâmica de transmissão de dados não pode ser descrita pela análise individual de seus roteadores. De maneira similar, a transmissão de sinais do cérebro é fundamentalmente afetada pela organização das sinapses entre neurônios (8). De fato, o estudo individual de tais células não é suficiente para se entender os processos cognitivos bem como muitas doenças, como a epilepsia (9). Logo, o estudo de sistemas complexos exige uma abordagem holística, que tem como princípio o fato de que o comportamento desses sistemas não pode ser explicado apenas pela soma de seus componentes.

Como um sistema complexo é representado por partes individuais que se conectam, é natural modelá-los como redes. Sendo assim, para analisar a estrutura e dinâmica de sistemas complexos, foi introduzida no final da década de 90 a teoria das redes complexas. Devido a essa generalização na representação de sistemas complexos, isto é, vértices conectados por arestas, a teoria das redes complexas vem sendo aplicada nas mais diversas áreas de pesquisa (10, 11). A motivação da modelagem por redes surgiu com a observação de que a organização de muitos sistemas complexos não é aleatória, mas está ligada às características particulares de cada sistema e leis que guiam sua evolução (12, 13). Por exemplo, no caso da Internet, os roteadores mais importantes, localizados em grandes centros urbanos, são os mais conectados na rede e tendem a receber mais conexões ao longo do tempo do que os roteadores localizados em cidades menores. De maneira semelhante, há certas pessoas na sociedade que possuem um número elevado de amigos, enquanto a maioria da população se organiza em pequenas comunidades — como se pode observar em sites de relacionamento como *Orkut* e *Myspace*. Nesses dois casos, a topologia da rede se auto-organiza de forma que haverá alguns vértices altamente conectados, enquanto que a maioria deles apresenta poucas conexões. Na área de biologia de sistemas, por exemplo, as interações proteicas, metabólicas ou genéticas são analisadas de uma maneira global, analisando-se a estrutura das conexões como um todo (14). De fato, esse tipo de análise vem se revelando fundamental para o desenvolvimento da microbiologia (14).

A representação por redes também tem sido usada com sucesso em sociologia, no estudo de propagação de epidemias ou rumores (15); neurociência, na análise da relação entre a estrutura do cérebro e funções cerebrais (8); engenharia, na determinação da vulnerabilidade de redes de transmissão de energia (16); linguística, no estudo das relações semânticas entre as palavras e em análises de autoria de textos (17); economia, na caracterização das relações entre empresas e transações comerciais (18); e meteorologia, na análise das relações climáticas entre diferentes pontos do globo (19)\*.

Nessa dissertação, estamos interessados na análise da vulnerabilidade de redes complexas. Diversos trabalhos têm revelado que redes complexas se auto-organizam de modo a garantir a robustez do sistema quando ocorrem falhas aleatórias, ou seja, quando vértices são removidos sem qualquer prioridade (5). Apesar dos avanços obtidos na análise de resiliência, diversas aplicações ainda não foram investigadas, como a comparação da resiliência de redes biológicas de diferentes organismos ou de malhas rodoviárias de diversos países. Sendo assim, nesse trabalho analisamos tais sistemas complexas e mostramos como essas redes podem ter sua robustez quantificada em termos de medidas de entropia. Além disso, verificamos quais propriedades estruturais mais contribuem para a robustez de redes biológicas e analisamos como a robustez

<sup>\*</sup> Uma descrição de todas as aplicações de redes é realizada na referência (10). Para maiores detalhes sobre o estudo de diversos fenômenos dinâmicos, ver a referência (20).

de uma malha rodoviária está ligada à facilidade de acesso em tal rede.

Os estudos aqui introduzidos devem contribuir para o entendimento da relação entre estrutura e dinâmica de sistemas complexos, o que é um tema de fundamental interesse da comunidade científica atualmente (11).

#### 1.1 Objetivos

Os objetivos básicos de nosso trabalho podem ser resumidos nos seguintes tópicos:

- Análise e caracterização da robustez de redes complexas através da simulação de falhas e ataques.
- Quantificação da robustez de redes através do uso de medidas de entropia da distribuição das conexões e entropia dinâmica relacionada à cadeias de Markov.
- Comparação do nível de resiliência de redes de interação proteína-proteína de quatro espécies, isto é, Saccharomyces cerevisiae, Caenorhabditis elegans, Drosophila melanogaster e Homo sapiens.
- Verificação de quais propriedades estruturais mais contribuem para a robustez de redes de interação de proteínas.
- Comparação de malhas rodoviárias de sete países, isto é, Brasil, Portugal, Polônia, Romênia, Austrália, Índia e África do Sul.
- 6. Introdução da medida de acessibilidade baseada em cadeias de Markov.
- 7. Análise da relação entre robustez e nível de acessibilidade em redes de rodovias.

#### 1.2 Contribuições

Dentre as contribuições principais deste trabalho, podemos citar:

- As redes de interação de proteínas de organismos com diferentes níveis de complexidade apresentam comportamento similar quando submetidas a falhas aleatórias, apresentando um alto grau de robustez.
- A verificação de que os organismos mais complexos, como homem e mosca, são mais robustos à ataques dirigidos do que os organismos mais primitivos, como a levedura.
- A determinação das propriedades estruturais que mais influenciam na resiliência de redes de proteínas indicam que o grau médio dos vizinhos e o coeficiente de escala são as propriedades que mais contribuem para a robustez.
- A comparação das malhas rodoviárias de diferentes países e sugere que os mais robustos são a Romênia, seguida de Portugal e Brasil.
- Introdução de uma generalização da medida de acessibilidade em redes.
- A análise em nível local permitiu mostrar que as cidades localizadas nas fronteiras dos países e no litoral são as que menos contribuem para a robustez de uma rede.
- A medida de acessibilidade introduzida apresenta alta correlação com a medida de entropia dinâmica, usada para quantificar o nível de resiliência de uma rede. Tais medidas apresentam pouca correlação com o grau de intermediação, que é uma medida de centralidade clássica. Sendo assim, a medida de acessibilidade quantifica a centralidade de vértices não em termos das menores distância, mas leva em conta todas as caminhadas possíveis em um processo de Markov.
- Nossos resultados podem ser utilizados no planejamento de novas rodovias, de modo que as novas ligações que forem estabelecidas devem priorizar as rodovias que contribuam

34
significantemente para o aumento dessas três propriedades de uma rede, isto é, robustez, acessibilidade e centralidade. Utilizando um método de reconhecimento de padrões, mostramos que tais cidades são aquelas localizadas nas fronteias e litoral dos países, pois essas apresentam pouco acesso.

### 1.3 Descrição dos capítulos

Iniciamos nosso estudo sobre redes com o Capítulo 2, onde veremos os conceitos fundamentais e a história do surgimento da ciência das redes. Faremos uma revisão dos tipos de redes, formas de representação e como calcular medidas básicas que podem descrever as relações de interconexões dos vértices, como o grau de um vértice, ou seu coeficiente de aglomeração. Serão vistos os modelos que foram desenvolvidos para geração de redes, através dos quais podem ser estudados diversos sistemas complexos. Finalizando a descrição deste capítulo, apresentamos alguns exemplos de redes reais e uma tabela de valores de suas medidas básicas.

No capítulo 3, veremos medidas mais elaboradas que podem dar detalhes importantes da estrutura topológica, e que nos permitirão analisar e classificar as redes. Serão apresentadas medidas que caracterizam a correlação entre o grau dos vértices, centralidade de um vértice, assim como, medidas de entropia e entropia dinâmica, as quais são utilizadas para descrever a resiliência de redes contra falhas e ataques. Será descrita uma dinâmica de ataque dos vértices de uma rede, onde pode se observar o comportamento da estrutura da rede sob perturbações, e assim quantificar sua robustez. Também será visto, que as medidas de entropia pode ser utilizadas como medidas de resiliência da rede.

Nos dois capítulos seguintes, foram examinadas redes reais com o intuito de quantificar suas características de robustez e acessibilidade. O Capítulo 4 mostra a análise de redes biológicas, onde estudamos as redes de interação de proteínas de quatro espécies, buscando uma relação entre a estrutura da rede e suas respectivas funções biológicas. Analisamos computacionalmente

a resiliência da estrutura de rede, sob uma dinâmica que simula ataques e falhas dos vértices (as proteínas). Várias medidas estruturais foram extraídas, a fim de classificar as redes de proteínas quanto aos modelos de redes estudados. Medidas dinâmicas foram calculadas para caracterizar a robustez das redes, e também diferenciá-las quanto a complexidade biológica, onde observamos a relação entre estrutura e função. As medidas de robustez (como as medidas de entropia) foram correlacionadas com medidas estruturais, para descobrir que medida estrutural estava mais fortemente colaborando para a resiliência da rede à falhas.

No Capítulo 5 analisamos as redes geográficas de sete países, e para estudar o fluxo de transporte de suas malhas rodoviárias, foi introduzido uma generalização da medida de acessibilidade, a qual é relacionada à facilidade de encontrar um certo vértice na rede ou, no caso, quanto uma cidade é acessível em um país. Apresentamos um método para rotular os cruzamentos de vias e propagar os rótulos através da malha, e assim foi possível construir a rede de interligação de cruzamentos. A entropia dinâmica foi usada novamente para quantificar a robustez de redes. Visando melhorar o fluxo de transportes para um determinado país, foram identificados os cruzamentos com menores valores de acessibilidade e robustez (para os quais deve-se priorizar o investimento em infra-estrutura rodoviária). Por fim, verificamos que existe uma forte relação entre a acessibilidade e a medida de entropia dinâmica, permitindo utilizar esta nova medida para caracterizar a robustez de redes.

No último capítulo, discutimos as conclusões e as possibilidades de trabalhos futuros, logo que a teoria das redes complexas mostra-se, cada vez mais, infindável para suas aplicações.

# 2 CONCEITOS BÁSICOS

### 2.1 O que é uma rede

O estudo de sistemas complexos ficou reconhecido pela sua interdisciplinariedade, devido aos avanços que se tem conseguido em diversos campos que vão da física à antropologia. Muitos sistemas ao nosso redor são ditos complexos, pois são formados por muitos elementos capazes de interagir entre si e com o meio ambiente (21, 22). Para se compreender o comportamento destes sistemas, nós devemos entender não somente o comportamento de suas partes mas como elas interagem para gerar o comportamento do sistema em nível global (22).

As principais propriedades de sistemas complexos são as seguintes: (i) *emergência*: a complexidade do todo é maior do que a complexidade da soma das partes, (ii) *auto-organização*: o sistema se organiza sem um comando externo e (iii) *universalidade*: sistemas pertencentes à mesma classe possuem propriedades semelhantes (22). Como esses sistemas são formados por partes discretas que se conectam, a modelagem por redes é realizada de forma natural (21, 22). Por exemplo, o cérebro é formado por células nervosas conectadas por axônios (2, 23); a sociedade é formada por pessoas ligadas por laços de amizade (24); as cadeias alimentares são formadas por animais e plantas conectados por relações de predador e presa (25); as conexões entre aeroportos são definidas por rotas aéreas (26, 27); as malhas rodoviárias são formadas por rodovias que ligam cidades (28); a Internet é composta por roteadores ligados por fibras-ópticas (5, 29, 30); a Teia Mundial é definida por documentos conectados por *hyperlinks* (31–33); a linguagem é resultante da interação de palavras que se relacionam por similaridade e função (34, 35) e os cientistas estão conectados de acordo com os trabalhos de colaboração (36, 37). Aplicações da teoria das redes complexas na modelagem de sistemas naturais e artificiais são encontradas em diversos artigos de revisão (21, 38–40) e livros (6, 41–43). Mas o que é exatamente uma rede? Quais os tipos de topologia mais comuns? Como pode-se caracterizar tais estruturas? Antes mesmo de iniciar a explicação a essas perguntas, veremos uma breve história do surgimento do conceito de redes complexas.

## 2.2 Introdução histórica

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. Já a teoria dos grafos se iniciou em 1736, quando o matemático e físico suíço Leonard Euler (44) se interessou por um problema de interligação de vértices, no qual a solução se resumia a simples resolução de um grafo, ou seja, descrever a rede de interligações apenas pela sua estrutura topológica. O problema que Euler acabou solucionando, ficou conhecido como o problema das setes pontes de Königsberg – uma cidade construída sobre as margens do rio Pregel, na antiga Prússia (hoje, a cidade pertence ao território russo e é chamada de Kalimingrad). Ao longo do rio existem duas ilhas que dividem o fluxo das águas em canais, e o acesso às ilhas era feito através de sete pontes, como pode ser visto pela Figura 2.1(a).

Na época, um quebra-cabeça para os habitantes, era tentar responder a seguinte pergunta: "Existe um caminho que atravessa todas as sete pontes, passando apenas uma única vez por cada uma delas?" As pessoas na cidade de Königsberg gastavam horas tentando solucionar este quebra-cabeça. Mas, por fim, Euler provou ser impossível a existência de tal caminho. Ele definiu as ilhas e as margens como simples pontos, ou vértices, e as pontes como arestas, ou ligações. Euler fez uso de um grafo, como da Figura 2.1(b), que representava muito facilmente o problema, descrevendo-o de uma forma matemática.



**Figura 2.1** – Diagrama em (a) das pontes da cidade de Königsberg (retirado de um artigo (44) publicado por Euler), e sua representação no grafo em (b).

A solução do problema das pontes de Königsberg parecia simples para Euler. Todos os vértices (ilhas ou margens) tinham um número *ímpar* de arestas (as pontes), mas ao fazermos um percurso pelo grafo, toda chegada a um vértice através de uma aresta e, logo após, uma correspondente saída do vértice através de outra aresta, sempre resultará em um número *par* de arestas percorridas, o que mostra ser impossível de existir um caminho que permitisse percorrer todas as pontes uma única vez. Sendo assim, o grafo não apresentava um caminho fechado contendo todas as arestas visitadas uma vez cada (caminho *Euleriano*). Muitos consideram esta prova realizada por Euler como o primeiro teorema desenvolvido no campo da matemática discreta, que contém a teoria dos grafos (45).

Somente depois de dois séculos, cientistas começaram demonstrar um crescente interesse na modelagem de grafos. Eles começaram a pesquisar os grafos como sendo o meio através do qual a informação, ou uma doença, pode ser difundida pela rede. De fato, cientistas como os húngaros Erdös e Rényi (46) e o russo Solomonoff (47), mostraram que, as propriedades estruturais de uma rede estão ligadas a processos dinâmicos como, por exemplo, a máxima propagação esperada de uma epidemia (47). Assim, no início da década de 60, iniciaram-se as pesquisas em redes aleatórias (48), e em 1967, uma importante propriedade presente nas redes sociais foi descoberta quando Stanley Milgram, um pesquisador de sociologia em Harvard, Estados Unidos, interessado na estrutura da sociedade americana, descobriu que a distância média entre duas pessoas quaisquer nos Estados Unidos é próxima de seis. Esse fenômeno é conhecido como efeito de mundo pequeno (6, 49).

Apesar dos avanços obtidos em sociologia, a falta de poder computacional e bases de dados de redes reais fizeram com que as investigações em redes complexas se reduzissem à análise de redes formadas por algumas dezenas ou no máximo centenas de vértices. Além disso, o estudo de redes sociais limitava-se a caracterizar a estrutura das ligações, como no trabalho de Milgram, e não em modelar as interações sociais ou investigar as propriedades dinâmicas dessas redes, como sua evolução. Apenas a partir do final da década de 90, com o advento da Internet e com o aumento do poder computacional, a teoria das redes complexas foi estabelecida, pois bases de dados foram disponibilizadas e os computadores tornaram-se capazes de processar um grande volume de informações.

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) 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. Além de princípios da teoria dos grafos, a teoria das redes complexas apresenta conceitos de física estatística, sistemas não lineares, fractais e autômatos celulares, dentre outras áreas. Esse caráter multidisciplinar permite a teoria das redes complexas cobrir aplicações desde biologia até sociologia, sendo a computação responsável pelas ferramentas utilizadas na modelagem, simulação e tratamento das bases de dados.

#### 2.3 Como representar uma rede

Representamos a estrutura de uma rede complexa através de um conjunto  $\mathscr{R}$  que, no caso de redes que não apresentam pesos em suas ligações, é definido por  $\mathscr{R}(\mathscr{V},\mathscr{E})$ , o qual é composto por um conjunto de *N* vértices,  $\mathscr{V} = \{v_1, v_2, \dots, v_N\}$ , e um conjunto de *M* conexões,



**Figura 2.2** – Representação em (a) de uma rede não dirigida de 6 vértices, e em (b) seu armazenamento computacional na forma de lista de conexões.

 $\mathscr{E} = \{e_1, e_2, \dots, e_M\}$ , que contém pares (i, j) de elementos de  $\mathscr{V}$ . O tamanho da rede está relacionado com número total de elementos do conjunto de vértices. As conexões podem ser dirigidas, quando o sentido da ligação importa, ou não-dirigidas.

Se as ligações possuem intensidade, como no caso da Internet que necessita diferenciar as arestas pela largura da banda disponível, a cada aresta é associado um peso. Neste caso, a rede deve apresentar informações adicionais sobre os pesos, isto é, além de ser formada pelos conjuntos  $\mathcal{V}, \mathcal{E}$ , a rede possui ainda o conjunto  $\mathcal{W} = \{w_1, w_2, \dots, w_M\}$ , que representa o peso das ligações, sendo a rede representada por  $\mathcal{R} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ . Para as redes não dirigidas, a representação das conexões é feita apenas por linhas, não diferenciando o sentido da ligação, ou seja, (i, j) e (j, i) representam a mesma conexão, vejamos em detalhes.

#### 2.3.1 Redes não dirigidas

Para uma rede não dirigida com *N* vértices, podemos rotulá-los com os números inteiros de um conjunto  $\mathscr{V} = \{1, ..., N\}$ , como o exemplo da rede na Figura 2.2(a), e listar as conexões por pares de elementos deste conjunto  $\mathscr{V}$ . Esta rede pode ser armazenada computacionalmente na forma de uma lista de conexões de pares de vértices *i* e *j*, como pode ser observado na Figura 2.2(b), ou na forma de uma *matriz de adjacência*  $\mathscr{A}$  (descrita pela matriz 2.2).

As listas de conexões são muitas vezes utilizadas para gravar a estrutura de redes em arquivos de computador, porém a forma mais prática de representar uma rede é através de uma matriz de adjacências  $\mathscr{A}$ , devido a sua aplicação na construção de algoritmos que se utilizam dessa matriz para calcular medidas topológicas. Os elementos  $a_{ij}$  desta matriz, são dados por:

$$a_{ij} = \begin{cases} 1 & \text{, se existir uma conexão entre os vértices } i \in j, \\ 0 & \text{, caso contrário.} \end{cases}$$
(2.1)

Deve-se notar que a matriz 2.2 apresenta na sua diagonal principal apenas elementos nulos, indicando a falta de auto-conexão que liga o vértice a ele mesmo. Também podemos observar que a matriz  $\mathscr{A}$  é simétrica, ou seja, o elemento  $a_{ij}$  tem o mesmo valor de  $a_{ji}$ , pois nestas redes não se diferencia o sentido da ligação.

$$\mathscr{A} = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \end{pmatrix}$$
(2.2)

## 2.3.2 Redes dirigidas

1

Também chamadas de dígrafos, as redes dirigidas são redes cujas conexões são dirigidas de um vértice *i* a um vértice *j*, sendo representada graficamente por uma seta. Este tipo de rede apresenta uma variação na definição dos elementos  $a_{ij}$  da matriz de adjacência, desta forma:

$$a_{ij} = \begin{cases} 1 & \text{, se existir uma conexão } do \text{ vértice } i \text{ ao } j, \\ 0 & \text{, caso contrário.} \end{cases}$$
(2.3)

Um exemplo de rede dirigida está representado na Figura 2.3, e foi baseado no exemplo



Figura 2.3 – Representação de uma rede dirigida de 6 vértices.

anterior, mas recebeu uma informação a mais no momento da leitura da lista de conexões (Figura 2.2(b)), informando que o sentido de formação da ligação é relevante. Para este tipo de rede, a matriz de adjacência 2.4 (denotada também por  $\mathscr{A}$ ) não é mais simétrica como pode-se observar.

#### 2.3.3 Redes com pesos

As conexões das redes foram até agora descritas apenas se existe a ligação (valor um) ou se não existe (valor zero), mas não é a única forma possível na construção das redes, inclusive muitas redes não são descritas desta forma binária (0 ou 1), mas sim pela utilização de pesos nas ligações, como a rede da Internet, onde os pesos estão relacionados com largura da banda de dados através das conexões, como também as cadeias alimentares, onde os pesos estão relacionados com o fluxo de energia das presas para os predadores. O exemplo de rede com pesos está representado na Figura 2.4(a), e a lista de conexões recebeu uma coluna a mais que identifica



Figura 2.4 – Representação em (a) de uma rede com pesos em suas ligações, e seu armazenamento na forma de lista de conexões em (b).

os pesos w das ligações, o que pode ser observado na Figura 2.4(b). Existem redes com pesos dirigidas ou não dirigidas, dependendo do tipo da rede complexa que se está estudando.

Neste caso, a matriz de pesos 2.5, representada por  $\mathcal{W}$ , apresenta um diferencial, em relação as já observadas, devido ao fato de conter números reais. Tais números representam os pesos das ligações e podem ser entendidos como a proporção de força da ligação, por exemplo, a conexão do vértice 2 ao vértice 4 tem peso 1, e é duas vezes mais forte que a conexão do vértice 1 ao vértice 2 que tem peso 0,5.

# 2.4 Medidas básicas para caracterização de redes

Nessa seção veremos quais as medidas mais comuns que podemos calcular para caracterizar a estrutura de redes complexas. Tais medidas são fundamentais, pois como a maioria das redes complexas são formadas por milhares (ou mesmo milhões) de vértices, apenas a sua visualização não é suficiente para obtermos informações sobre a sua estrutura, conforme notamos na Figura 2.12 (exemplos de redes reais).

#### 2.4.1 Conectividade

Uma medida muito importante de redes é chamada conectividade, ou grau, porém simples de ser calculada. Conectividade e grau são sinônimos, no caso de não existir multi-conexões entre os pares de vértices, e apenas este caso será tratado nesta dissertação. O grau de um vértice *i* é dado pelo número de arestas conectadas a ele (no caso de redes não dirigidas) e que denotamos por  $k_i$ , o qual pode ser calculado pela equação a seguir, em termos da matriz de adjacência, onde  $a_{ij}$  é o elemento (i,j) da matriz  $\mathscr{A}$ ,

$$k_i = \sum_{j=1}^{N} a_{ij}.$$
 (2.6)

Como o grau de um vértice é uma medida local, para caracterizar a estrutura de redes, temos um medida global que é calculada obtendo-se a média amostral do número de conexões entre os vértices, a qual chamamos de conectividade média ou grau médio  $\langle k \rangle$ ,

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^{N} k_i. \tag{2.7}$$

#### 2.4.2 Distribuição de conectividade

A organização das conexões de uma rede complexa pode ser caracterizada através da distribuição das conexões. Denotamos esta distribuição por P(k), onde k é a conectividade, e o valor de P(k = d) representa a probabilidade em que pode ser encontrado um vértice com k = d conexões na rede.



**Figura 2.5** – (a) Rede social de amizades em um clube de karate (1). (b) Histograma da distribuição de conectividade normalizada.

A construção da distribuição de conectividade é feita através de um histograma normalizado da quantidade de vértices para todos valores de k, conforme a Figura 2.5(b), onde utilizamos a rede social de amizades em um clube de karate (1) da Figura 2.5(a).

#### 2.4.3 Coeficiente de aglomeração

Redes complexas geralmente apresentam ciclos de ordem três, que são ciclos de três vértices conectados, como a rede neural do *Caenorhabditis elegans* e a rede de colaboração de atores (2). Uma maneira de caracterizar a presença de laços (*loops*) desta ordem é através do coeficiente de aglomeração (*clustering coefficient*), o qual denominamos por *cc*. Para redes sem pesos, o coeficiente de aglomeração *cc<sub>i</sub>* (equação 2.8) para um vértice *i*, é definido como a razão entre o número de ligações *e<sub>i</sub>* entre os nós da sua vizinhança e o número total de arestas que possivelmente poderia existir entre eles, isto é,  $(k_i(k_i - 1)/2)$ . Este último termo pode ser entendido como o produto da quantidade de vizinhos  $(k_i)$  pela conectividade máxima  $(k_i - 1)$  de cada um destes (devido ao fato que uma aresta é compartilhada por um par de vizinhos, torna-se necessário a divisão por dois).

Esta relação que define o coeficiente de aglomeração pode apresentar valores entre 0, no caso dos vizinhos não se conectarem entre si, e no máximo 1, para o caso de todas as possíveis conexões entre os vizinhos terem sido preenchidas, como mostra a Figura 2.6.



Figura 2.6 - Exemplificação para os valores de coeficiente de aglomeração calculados para o vértice em vermelho.

Uma outra medida global para quantificar a ocorrência de triângulos em redes é dada pela média do coeficiente de aglomeração para todos os vértices, o qual denominamos por  $\langle cc \rangle$ , e que é calculado pela seguinte equação,

$$\langle cc \rangle = \frac{1}{N} \sum_{i=1}^{N} cc_i.$$
(2.9)

### 2.4.4 Distâncias

Um *caminho* em redes é uma sequência de vértices  $i_1, i_2, ..., i_n$  tal que para cada um destes vértices exista uma aresta para o próximo vértice na sequência. O comprimento de um caminho ou *distância* entre  $i_1$  e  $i_n$  é portanto o número de arestas deste caminho, que é igual a n - 1, para redes sem pesos. Para redes com pesos, o comprimento é a adição de cada peso em cada aresta.

O transporte ou comunicação em redes está relacionado com a distância entre os vértices. Por exemplo, em redes geográficas como as redes de rodovias (Figura 2.7), o caminho entre dois pontos está relacionado com o tempo de transporte e a ocorrência de congestionamento (50). As distâncias também influenciam na transmissão de dados na Internet, onde o atraso ou a perda de pacotes trocados está relacionada as rotas, ou caminhos tomados, podendo ocorrer falhas e



Figura 2.7 – Rede geográfica de rodovias do Brasil. Os pontos representam as cidades.

congestionamentos de roteadores (51), os quais se assemelham aos entroncamentos de rodovias. Por isso, os caminhos e suas distâncias são fundamentais para o estudo de redes reais.

O comprimento de um caminho é medido pelo número de arestas existentes entre o par de vértices (i, j). Podemos definir o *comprimento do menor caminho*, como sendo  $d_{ij}$ , selecionandose o menor dentre todos caminhos possíveis ligando  $i \in j$ . Os menores caminhos entre todos os pares de vértices formam um conjunto representado na matriz de distâncias D, contendo nos seus elementos  $d_{ij}$  a menor distância entre os pares  $i \in j$ . O máximo valor encontrado nesta matriz corresponde ao *diâmetro* da rede, denominado por  $d_{max}$ , enquanto que a média sobre todos os valores da matriz D retorna o *menor caminho médio*  $\ell$ , calculado por

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

onde levamos em consideração apenas componentes conectados, pois no caso de *i* e *j* não pertencerem a um mesmo componente, temos  $d_{ij} = \infty$ , e o somatório diverge.

# 2.5 Modelos básicos

Além de medidas para caracterização topológicas, pesquisadores de teoria de redes complexas têm desenvolvido modelos de redes para predizer a evolução da organização de sistemas complexos. Tais modelos permitem que se execute basicamente quatro tarefas: (i) predição, que permite descrever como a estrutura do sistema evolui no tempo; (ii) simulação, que auxilia na análise de diversos processos dinâmicos, como fluxo e propagação de epidemias, que podem ser simulados em topologias que representem uma rede complexa, como a Internet e a sociedade; (iii) controle de processos, sendo possível a reprodução de condições correspondentes à realidade em ambientes virtuais e o monitoramento do comportamento do sistema quando sujeito a algum processo dinâmico, como no caso da modificação da largura de banda na Internet; e (iv) estudo de falhas, onde é possível verificar-se os efeitos de remoções de conexões e vértices em uma rede, como ocorre no caso da Internet, onde roteadores podem falhar, gerando redistribuição do tráfego e levando a perda de desempenho. Assim, com a modelagem, o controle, previsão e o desenvolvimento de mecanismos para tornar sistemas complexos mais robustos podem ser realizados de maneira eficiente a baixo custo em ambientes virtuais.

Dezenas de modelos redes complexas têm sido desenvolvidos para o estudo dos mais diversos sistemas complexos (20). Dentre esses modelos, os mais clássicos e, geralmente usados, são descritos a seguir.

#### 2.5.1 Redes aleatórias

Em 1959, dois matemáticos húngaros, Paul Erdős e Alfred Rényi, consideraram os grafos como objetos estocásticos, ao invés de analisá-los de forma puramente determinística, como fazia até então a matemática discreta e a sociologia. Segundo tal modelo, que denominaremos apenas por ER, redes são construídas iniciando-se com um conjunto de *N* vértices totalmente



**Figura 2.8** – (a) Exemplo de uma rede aleatória de Erdös e Rényi. (b) Distribuição da conectividade para uma rede com N=1000, usando uma probabilidade p = 0, 2. Os pontos no gráfico são a média de 100 simulações.

desconectados. A seguir, a cada passo do algoritmo, dois vértices são escolhidos aleatoriamente e conectados com uma probabilidade fixa *p*, sendo cada par de vértices considerado apenas uma vez. Assim sendo, todas as ligações possuem a mesma probabilidade de ocorrerem, ou seja, a rede gerada tem uma estrutura altamente homogênea.

A distribuição de conectividade P(k) possui um valor característico dado pela conectividade média  $\langle k \rangle = p(N-1)$ , logo que a probabilidade p defini a proporção de candidatos a formarem conexões. Para uma valor alto de N e baixo de p, esta distribuição se aproxima a uma distribuição de Poisson (Figura 2.8(b)), devido à lei dos eventos raros para um processo de Bernoulli (52), isto é,

$$p(k) = \lim_{N \to \infty, p \to 0} \frac{N!}{(N-k)!k!} p^k (1-p)^k \simeq \frac{e^{-\langle k \rangle} \langle k \rangle^k}{k!}, \qquad (2.11)$$

onde  $\langle k \rangle = p(N-1)$  é o grau médio da rede.

Diferentemente da grade regular, no modelo *ER* ocorre uma quebra de simetria que faz com que a distância média  $\ell$  entre os vértices (equação 2.10) seja pequena se comparada ao tamanho da rede, que é da ordem de  $\ln(N)/\ln(\langle k \rangle)$ , e esta distância diminui conforme a rede se torna esparsa tendo *N* muito maior que o grau médio.

#### 2.5.2 Modelo de Watts-Strogatz

Em 1998, analisando dados empíricos, pesquisadores das Universidade de Columbia e Cornell observaram que em algumas redes reais, tais como a rede de neurônios do *Caernohabditis elegans* e a rede de distribuição de energia dos Estados Unidos; a presença de *loops* (caminhos fechados) de ordem três é muito maior do que nas redes aleatórias com mesmo número de vértices e arestas (2). Esse foi o primeiro indício de que as redes reais não são completamente aleatórias, mas possuem leis que regem sua organização. Baseados nesta descoberta, Watts e Strogatz sugeriram um modelo alternativo aos grafos aleatórios, chamado modelo *small-world de Watts-Strogatz* (em analogia ao fenômeno descoberto por Stanley Milgram).



**Figura 2.9** – Exemplificação da variação da probabilidade p no modelo de Watts-Strogatz com conectividade média  $\langle k \rangle = 4$ .

Nesse modelo, que denominaremos por WS, para a construção de um rede, inicia-se com uma grade regular formada por N vértices ligados a  $\kappa$  vizinhos mais próximos em cada direção, totalizando  $2\kappa$  conexões iniciais, sendo  $N \gg \kappa \gg \log(N) \gg 1$ . A seguir, cada aresta é aleatoriamente reconectada com uma probabilidade fixa p, que introduz o caráter aleatório à rede. Quando p = 0 a rede é completamente regular e quando p = 1, a rede é aleatória. Portanto, tal modelo se situa entre a completa regularidade e a aleatoriedade. A emergência do regime *small-world* ocorre para p > 0,01 (2). Tal fato pode ser observado na Figura 2.10(a), a variação do coeficiente de aglomeração cc e do caminho médio  $\ell$  entre os vértices, em relação aos valores iniciais quando p = 0. Apenas para valores de p > 0,01 é que podemos observar o efeito mundo pequeno na geração destas redes, pois abaixo deste valor a rede se aproxima da grade



**Figura 2.10** – (a) Gráfico do coeficiente de aglomeração *cc* e do caminho médio  $\ell$  para os diferentes valores de probabilidade *p* na geração de redes do modelo *WS* (2). (b) Distribuição da conectividade para uma rede com N=1000,  $\langle k \rangle = 200$ , usando uma probabilidade p = 0,01. Os pontos no gráfico são a média de 100 simulações.

regular, e acima deste valor temos uma queda rápida de *cc* e também valores muito baixos de  $\ell$ . Isto se deve ao rearranjo das conexões, pois vértices que antes não eram vizinhos podem agora compartilhar novas ligações, o que diminui a distância entre eles e seus vizinhos próximos. A distribuição de conectividade destas redes é mostrada na Figura 2.10(b), sendo que os vértices apresentam um número típico de conexões, como ocorre no model ER.

#### 2.5.3 Redes livres de escala

Apesar de redes sociais apresentarem algumas propriedades das redes aleatórias como, por exemplo, o efeito mundo pequeno, conforme visto anteriormente, estas redes não apresentam uma distribuição de conectividade de Poisson. Logo, os modelos ER e WS não são adequados para representar a evolução e organização dessas redes.

Em 1999, Albert-László Barabási e Réka Albert (13), decidiram analisar a topologia da Teia Mundial (*World Wide Web*) usando um *web crawler*. Desse modo, eles mapearam a topologia das conexões entre as páginas da Teia Mundial e descobriram que além dela apresentar o fenômeno *small-world* ( $\ell \approx 11$ ), a distribuição de conexões não é aleatória, mas do tipo *livre de*  escala (scale-free), que é da forma  $P(k) \sim k^{-\gamma}$  (31). Essa descoberta causou um certo espanto, pois era esperado que a Teia Mundial fosse organizada de forma aleatória, pois essa não possui um comando central para definir suas conexões. No entanto, naquele mesmo ano, o mesmo tipo de distribuição já havia sido encontrada na Internet pelos irmãos Faloutsos (29), embora tal trabalho não tivesse despertado forte impacto. Ou seja, diversas evidências sugeriram que a organização da maioria dos sistemas complexos não é aleatória, mas ocorre segundo ligação preferencial, conforme observado posteriormente, em redes de colaboração entre cientistas (36), redes metabólicas de 43 organismos diferentes (53) e redes de interação de proteínas (54), sendo que nesse último caso verificou-se que o número de conexões está relacionado com a letalidade de proteínas. Ou seja, a distribuição das conexões está relacionada às funcionalidades de um sistema complexo.

Como diversos sistemas complexos apresenta conexões do tipo livre de escala, o universo aleatório de Erdős e Rényi tende a não estar presente na natureza. Sendo assim, o trabalho de Barabási e Albert descartou a aleatoriedade e mostrou que há leis que regem a estrutura das redes reais. A distribuição livre de escala é um tipo de distribuição de probabilidades que reflete invariância de escala. Leis de potência também são conhecidas como Lei de Zipf, distribuição de Pareto ou lei de potência. Tais distribuições são idênticas, embora alguns autores tenham gerado confusão ao diferenciá-las (55). A distribuição de Pareto foi proposta por Vilfredo Pareto (56) no início do século 20, que demonstrou que certos fenômenos em economia, assim como na física, podem ser modelados matematicamente. Ele observou que em muitos fenômenos físicos e econômicos 80% das consequências eram devido aos 20% das causas, o que ficou conhecido como a regra 80/20, que gera uma uma curva decrescente e sem um pico característico, descrita por um único expoente negativo, o que chamamos de lei de potência ou livre de escala. Deste modo, a economia não é governada por simples aleatoriedade, mas possui leis que regem o seu comportamento. De fato, o tamanho de cidades, o diâmetro de crateras lunares, as distribuições de riquezas e o número de citações por artigo, seguem uma distribuição tipo lei de potência, apresentando a ausência de uma média e desvio que caracterizam uma escala típica (como em fenômenos aleatórios, onde quantidades geradas aleatoriamente possuem



Figura 2.11 – (a) Exemplo de uma rede livre de escala gerada pelo modelo SF. (b) Gráfico em escala logarítmica da distribuição de conectividade (em azul) e da distribuição cumulativa (em vermelho) para uma rede com N=1000. Os pontos no gráfico são a média de 100 simulações.

uma escala típica, veja a Figura 2.8(b)).

Motivados por essa ausência de escala em redes reais, Babrabási e Albert propuseram um modelo de crescimento, que gera redes livres de escala (13), que denominaremos por modelo SF, que é baseado em dois passos:

- Crescimento: Iniciando-se com um pequeno número de vértices N<sub>0</sub>, a cada passo é adicionado um vértice com m (m ≤ N<sub>0</sub>) arestas que se conectam com vértices já presentes na rede.
- Ligação preferencial: O novo vértice, que vai ser adicionado à rede, tende a se conectar com os vértices mais conectados, ou seja, a probabilidade de um vértice *j*, presente na rede, ser escolhido é proporcional a sua conectividade,

$$\mathscr{P}_{i \to j}(n+1) = \frac{k_j(n)}{\sum_{u=-N_0+1}^n k_u(n)},$$
(2.12)

onde *n* é o tempo e o número de vértices adicionados à rede.

Interessante notar que estes dois mecanismos de construção das redes livres de escala não estão presentes no modelo aleatório de Erdős e Rényi e no modelo *small-world* de Watts e

Strogatz, pois nestes não há adição de novos vértices e as conexões são estabelecidas de forma homogênea, havendo uma conectividade característica (conectividade média) na rede. Por outro lado, nas redes geradas pelo modelo livre de escala, os vértices mais conectados tendem a receber mais conexões. Este paradigma é conhecido como "*o rico fica mais rico*". Estas redes são formadas por um reduzido número de vértices altamente conectados, denominados *hubs*, e por uma grande quantidade de vértices pouco conectados, o que define a distribuição livre de escala. Na Figura 2.11 é mostrada uma rede gerada pelo modelo de Barabási e Albert e a distribuição da conectividade para uma rede composta por 1.000 vértices, assim como a sua distribuição cumulativa dada por  $P_{cum}(k) = \int_k^{\infty} P(x) dx$ .

#### 2.6 Exemplos de redes reais

As distribuições livre de escala têm sido observadas desde a década de 50 (57), mas sua popularidade cresceu com os trabalhos de Barabási e Albert no final da década de 90, pois foi apenas a partir dessa época que diversas bases de dados dos mais diversos sistemas complexos tornaram-se disponíveis. Deste modo, o desenvolvimento da Internet e a disponibilidade de informações que possibilitassem a construção de redes biológicas, sociais, tecnológicas e de informação, foram fundamentais para o desenvolvimento da teoria das redes complexas.

Apresentamos alguns exemplos de redes reais na tabela 2.1, onde é possível observar a variação no número de vértices N, conectividade média  $\langle k \rangle$ , menor caminho médio  $\ell$ , como também o expoente  $\gamma$  da distribuição de conectividade. A Figura 2.12 apresenta alguns exemplos de redes reais, como a rede da *World Wide Web*, a rede de conexões da Internet, a rede social de personagens do livro "Os miseráveis", de Victor Hugo, e a rede de interações de proteínas da levedura *Saccharomyces cerevisiae*, que será estudada no Capítulo 4.



**Figura 2.12** – Exemplos de redes complexas: (a) uma amostra da *World Wide Web*, (b) a Internet, (c) a rede social dos personagens do livro "Os miseráveis", de Victor Hugo, (d) a rede de interações de proteínas da levedura *Saccharomyces cerevisiae*.

Tabela 2.1	– Exemplos	de redes	complexas	reais o	que s	seguem	uma	distribuição	livre	de esc	ala
	· r · ·		r		1	0		5			

Rede complexa	Ν	$\langle k  angle$	l	γ	Referências
Atores	449.913	113,43	3,48	2,3	(2, 58)
Chamadas telefônicas	$47 \times 10^{6}$	3,16	-	2, 1	(59)
Mensagens de e-mail	59.912	1,44	4,95	1, 5 - 2, 0	(60)
Contatos sexuais	2.810	-	-	3,2	(61, 62)
WWW (nd.edu)	269.504	5,55	11,27	2, 1 - 2, 4	(31, 32)
WWW (Altavista)	203.549.046	10,46	16,18	2, 1 - 2, 7	(63)
Rede de citações	783.339	8,57	_	3.0	(64)
Co-ocorrência de palavras	460.902	70,13	-	2,7	(65, 66)
Internet	10.697	5,98	3,31	2,5	(12, 67)
Pacotes de softwares	1.439	1,20	2,42	1, 6 - 1, 4	(68)
Circuitos eletrônicos	24,097	4,34	11,05	3,0	(69)
Redes metabólicas	765	9,64	2,56	2,2	(53)
Interações de proteínas	2.115	2, 12	6,80	2,4	(54)

# *3 CARACTERIZAÇÃO DE REDES COMPLEXAS*

Para se entender o funcionamento de um sistema complexo, é fundamental conhecer-se sua organização. Redes complexas podem apresentar diferentes topologias, dependendo dos mecanismos que determinam sua evolução. Por exemplo, as redes sociais apresentam arquitetura muito diferente das redes tecnológicas, como a Internet, ou biológicas, como as redes de interação de proteínas (70). Para se quantificar a estrutura das ligações, diversas medidas têm sido desenvolvidas. Através de medidas, redes podem ser analisadas, caracterizadas, classificadas e modeladas (71). De fato, a análise e caracterização de redes é importante no estudo das relações entre forma e função em redes complexas. Por exemplo, se uma rede apresenta topologia do tipo livre de escala, é esperado que sua estrutura seja altamente tolerante à falhas. Ademais, a caracterização da estrutura das ligações de redes é importante na análise de simulações de processos dinâmicos sendo possível determinar quais aspectos topológicos influenciam na simulação de tais processos. Nessa seção, introduzimos diversas medidas de redes complexas que permitem caracterizar diferentes propriedades topológicas e dinâmicas de redes, tais como as medidas de assortatividade e entropia dinâmica, respectivamente.

### 3.1 Caracterização topológica de redes

Pesquisadores de redes complexas inicialmente estavam interessados no desenvolvimento de ferramentas para descrever a estrutura de redes, de modo a entender seu funcionamento e evolução (72). Desse modo, diversas medidas foram desenvolvidas de modo a permitir uma detalhada descrição da organização de sistemas complexos representados por redes.

#### 3.1.1 Assortatividade

Um grande número de redes reais são correlacionadas no sentido que a probabilidade de um vértice com grau k esteja conectado com um outro vértice com grau k' dependa de k (20). Nesse caso, pode-se utilizar o conceito de probabilidade condicional P(k'|k), sendo definida como a probabilidade que uma aresta de um nó com grau k se conecte com um nó de grau k'. Desse modo, a probabilidade condicional é dada por:

$$P(k'|k) = \frac{k'P(k')}{\langle k \rangle}$$
(3.1)

Observe que  $\sum_{k'} P(k'|k) = 1$ . Para redes não-dirigidas, P(k|k') = P(k'|k) e k'P(k|k')P(k') = kP(k'|k)P(k). Note que P(k|k') e P(k'|k) caracterizam a correlação entre o grau dos vértices (71). Embora a correlação seja formalmente definida por P(k'|k), podemos usar a média do grau dos vizinhos mais próximos de vértices com grau k, denotada por  $\langle k_{nn} \rangle$ , expressa da seguinte forma:

$$\langle k_{nn} \rangle = \sum_{k'} k' P(k'|k). \tag{3.2}$$

Quando  $\langle k_{nn} \rangle$  é uma função crescente em k, vértices com alto grau tendem a se conectarem

Rede	Ν	S	Referências
Colaboração em física	52.909	0,363	(37)
Colaboração em biologia	1.520.251	0,127	(37)
Colaboração em matemática	253.339	0,120	(73)
Colaboração de atores	449.913	0,208	(2)
Diretores de empresas	7.673	0,276	(74)
Internet	10.697	-0,189	(67)
World Wide Web	269.504	-0,065	(13)
Interação de proteínas	2.115	-0,156	(54)
Rede neural	307	-0,163	(2)
Cadeias alimentares	92	-0,276	(75)

Tabela 3.1 – Assortatividade em redes reais.



**Figura 3.1** – Figura adaptada do artigo (3). Média do grau dos vizinhos mais próximos de vértices com grau *k*, para a rede da Internet para os anos de 1997 a 1999.

entre si, estas redes são classificadas como assortativas, e caso  $\langle k_{nn} \rangle$  for decrescente em k, vértices com alto grau tendem a se conectarem com vértices que possuam baixo grau, elas são classificadas como disassortativas. Caso não ocorra correlação,  $\langle k_{nn} \rangle$  é independente de k (76).

Na Figura 3.1, observamos a relação da medida  $\langle k_{nn} \rangle$  com a conectividade *k* da Internet, seguindo uma dependência do tipo  $\langle k_{nn} \rangle \sim k^{-\nu}$ , para  $\nu = 0, 5$ . Este resultado indica que vértices altamente conectados são mais prováveis de se ligarem a vértices menos conectados, isto é, a rede da Internet é do tipo disassortativa.

Outra maneira de determinar a correlação de grau é considerando o coeficiente de correlação de Pearson nos extremos de uma aresta (76):

$$s = \frac{\sum_{ij} (a_{ij} - k_i k_j/2m) k_i k_j}{\sum_{ij} (k_i \delta_{ij} - k_i k_j/2m) k_i k_j}$$
(3.3)

onde  $a_{ij}$  é o elemento da matriz de adjacência  $\mathscr{A}$ ,  $k_i$  é a conectividade, m é o número de arestas e  $\delta_{ij}$  é a função delta de Kronecker, isto é,  $\delta_{ij} = 1 \operatorname{caso} i = j$ , e igual a 0 caso contrário. Se s > 0a rede é dita assortativa, se s < 0, a rede é disassortativa e para s = 0 não há correlação entre o grau dos vértices.

Enquanto que, em redes assortativas os vértices com graus similares tendem a se conectarem uns aos outros, em redes disassortativas os vértices altamente conectados tendem a serem ligados a vértices com baixa conectividade (76). Na tabela 3.1, podemos observar a que algumas redes reais são assortativas, como a rede de colaboração em física, enquanto que outras são disassortativas, como a rede de interação de proteínas.

#### 3.1.2 Conectividade média dos vizinhos

Além de calcular a média do grau dos vizinhos dependendo do grau k, também podemos calcular a conectividade média dos vizinhos de um vértice i, que definiremos por r(i) (20), como segue,

$$r(i) = \frac{1}{k_i} \sum_{j=1}^{N} a_{ij} k_j, \qquad (3.4)$$

onde  $a_{ij}$  é o elemento da matriz de adjacência  $\mathscr{A}$ , e  $k_j$  é a conectividade do vértice vizinho j. Podemos definir uma medida global para caracterização estrutural da rede, r, a qual é dada pela média amostral do grau médio dos vizinhos entre os vértices da rede, isto é,

$$r = \frac{1}{N} \sum_{i}^{N} r(i). \tag{3.5}$$

#### 3.1.3 Grau de intermediação

O grau de intermediação (*betweenness centrality*, em inglês) é uma medida que mede o quanto um dado vértice (ou aresta) *u* encontra-se nos caminhos entre outros vértices (77). Ou seja, essa medida quantifica o tráfego que passa por *u*, assumindo que este é feito pelos menores caminhos. Por exemplo, quando mensagens são trocadas através de uma rede social, podemos supor que cada par de vértices (pessoas) trocam mensagens com igual probabilidade por unidade de tempo, e que as mensagens sempre tomam os caminhos mais curtos. Então, para um tempo muito longo, o número de vezes que mensagens passam através de cada vértice será proporcional ao número de caminhos mais curtos que encontra-se o vértice. O grau de

intermediação é definido matematicamente por,

$$B_u = \sum_{ij} \frac{\sigma(i, u, j)}{\sigma(i, j)},\tag{3.6}$$

onde  $\sigma(i, u, j)$  é o número de menores caminhos entre os vértices *i* e *j* que passam pelo vértice (ou aresta) *u* e  $\sigma(i, j)$  é o número total de menores caminhos entre *i* e *j*. A soma é feita sobre todos os pares distintos *i*, *j* de vértices. Uma possível medida global para caracterizar uma rede é dada pela média do grau de intermediação para todos os vértices (ou arestas) *u*, isto é,

$$B = \frac{1}{N} \sum_{u}^{N} B_{u}.$$
(3.7)

Vértices com alto grau de intermediação podem ter grande influência dentro da rede, devido ao fato que eles têm controle sobre as informações que passam entre os demais vértices, e a remoção deles pode atrapalhar as comunicações entre os vértices.

#### 3.1.4 Entropia da distribuição das conexões

A entropia tem duas origens, uma das quais encontra-se em termodinâmica na física onde é usada como medida para ordem/caos e para a predição de estados de um sistema físico, chamada de entropia de Boltzmann. Esse conceito foi também introduzido por Shannon na área da teoria da informação, onde é usada como medida para a predição de uma fonte de informação. De fato, em uma discussão, Von Neumann (78) alertou Shannon para o fato que ambos os conceitos são matematicamente idênticos, então Shannon também nomeou sua medida de *entropia*. Para exemplificar, uma fonte de informação como uma cadeia de caracteres ou *string*, terá máxima entropia no sentido da entropia de Shannon, se gerar todos símbolos e padrões de símbolos com a mesma probabilidade. Podemos definir a medida H(X) como a incerteza na predição do próximo evento de uma variável aleatória discreta X, para a qual somente as probabilidades de eventos é definida

da seguinte forma:

$$H(p_1,...,p_n) = -\sum_{i=1}^n p_i \log p_i.$$
 (3.8)

Na teoria da informação, a unidade típica de medida é o *bit*, se a base do logaritmo é 2, ou *nat*, se o logaritmo natural for usado. As propriedades de *H* são:

- 1. H = 0, se todos, exceto um dos  $p_i$ , são zeros.
- 2. *H* é máxima, se  $p_1 = p_2 = \cdots = p_n$ .
- 3. caso  $p_i = 0$ , temos  $p \log p = 0$ , de acordo com  $\lim_{p \to 0+} p \log p = 0$ .

Em redes complexas, a entropia de Shannon é usada geralmente na análise da organização das conexões, onde leva-se em conta a distribuição do grau. A entropia *H* da distribuição de conectividade P(k), é definida por

$$H = -\sum_{k} P(k) \log[P(k)].$$
(3.9)

Esta medida quantifica a heterogeneidade na organização das conexões de uma rede complexa (80). Na Figura 3.2, podemos observar que a medida de entropia da distribuição de conectividade está relacionada com a robustez de redes. Nesse caso, o modelo *livre de escala* apresenta o maior valor de entropia e também uma maior resiliência à remoção aleatória de vértices, pois apresenta a menor variação do caminho médio entre os vértices l, em relação as outras duas redes. Este valor de entropia elevado para as redes livres de escala é devido ao fato da rede possuir uma heterogeneidade na distribuição das conexões (81), diferente das redes aleatórias que possuem uma distribuição mais homogênea.

### 3.2 Caracterização da robustez de redes complexas

Diversos tipos de informação podem se propagar por uma rede complexas, como a troca de pacotes entre computadores na Internet, ou mesmo, doenças sendo disseminadas entre in-



Figura 3.2 – Figura adaptada do artigo (4). Medida da varição do caminho médio entre os vértices ℓ em relação a remoção aleatória dos nós. Os valores da entropia da distribuição de conectividade também são indicados.

divíduos em uma sociedade. Para se analisar a propagação da informação, é necessário conhecerse a robustez da rede analisada sob o ponto de vista de falhas aleatórias e ataques, pois a propagação pode ser interrompida dependendo de quais vértices estão inacessíveis.

Tolerância à falhas é definida como a habilidade do sistema manter sua propriedades de comunicação após a remoção de uma fração f de seus vértices. Sistemas complexos são altamente robustos à falhas aleatórias, como pode observar-se na Internet, onde roteadores são desligados a todo instante, mas a transmissão de dados em nível mundial é mantida. Para se estudar a tolerância de redes complexas, devemos analisar sua organização quando vértices ou arestas são removidos. A Internet é um exemplo de rede altamente robusta, pois possui uma diversidade de caminhos, ou rotas, pelos quais a informação pode ir de um computador a outro, ocorrendo uma alta taxa de sucesso na entrega de pacotes entre roteadores. Mesmo se ocorrer uma falha em um roteador (*hub* da rede), o sistema da Internet é capaz de reorientar o caminho pelo qual a informação irá trafegar até alcançar seu destino correto. Embora falhas em roteadores seja muito comuns, ao se tratar a rede no nível mundial, a Internet se mantém em completo funcionamento, mesmo com esse nível de falhas. Em trabalhos recentes (5, 82) têm-se levantado a seguinte questão: se uma certa porcentagem de vértices falham, qual o efeito que teremos nas conexões entre os outros vértices da rede? Ou no caso da propagação de epidemias, se uma certa porcentagem de indivíduos são removidos da rede, por exemplo, sendo imunizados contra a doença, qual o efeito teremos na propagação (83) de tal doença? Seguindo estas ideias, iremos estudar os efeitos relacionados a falhas na estrutura de redes complexas.

#### **3.2.1** Falhas e ataques

A robustez de redes complexas pode ser analisada em termos de dois tipos de dinâmicas, *falhas* aleatórias ou *ataque*. No primeiro caso, vértices são removidos sem qualquer preferência, isto é, todos os vértices apresentam a mesma probabilidade de serem removidos. Por outro lado, os vértices são removidos de acordo com o seu número de conexões, desde os mais conectados aos menos conectados. Simulando tal remoção, podemos medir a variação do diâmetro da rede (Figura 3.3(a)), como também o tamanho da maior componente (isto é, o maior sub-grafo conectado) e o tamanho médio dos componentes isolados (Figura 3.3(b)), conforme ocorre a remoção de vértices. Desse modo, podemos observar os efeitos do aumento das distâncias entre os vértices e a diminuição do tamanho do maior componente, uma vez que a rede pode se fragmentar em sub-grafos.

O trabalho pioneiro que analisou a dinâmica falhas e ataques em redes complexas foi pu-



Figura 3.3 – Simulação de falhas e ataques na Internet: (a) diâmetro da rede, (b) tamanho da maior componente (símbolos abertos) e tamanho médio dos componentes isolados (símbolos fechados). Figuras adaptadas do artigo de Albert (5).



**Figura 3.4** – Simulação de falhas aleatórias e ataques em redes geradas pelos modelos livre de escala (SF) e aleatória (ER): (a) diâmetro da rede, (b) tamanho da maior componente (símbolos abertos) e tamanho médio dos componentes isolados (símbolos fechados). Figura adaptada do artigo de Albert (5)

blicado no ano 2000, por pesquisadores da Universidade Notre Dame (5). Nesse trabalho, os pesquisadores analisam como a estrutura de uma rede está relacionada com a robustez. Os autores analisaram a variação do diâmetro e do maior componente da Internet quando uma fração *f* de vértices é removida. Na Figura 3.3 são apresentados os resultados obtidos, onde observamos que a Internet é altamente robusta quando sujeita à falhas aleatórias, mas altamente vulnerável á ataques. A resiliência à falhas aleatórias se deve ao fato da seleção do vértice a ser removido ocorrer de forma aleatória, sendo baixa a probabilidade de se selecionar um vértice altamente conectado, pois a rede é do tipo livre de escala. Por outro lado, quando são simulados ataques, vértices altamente conectados são removidos e consequentemente ocorre uma aumento no diâmetro da rede, já que os vértices mais conectados são os responsáveis pelo efeito mundo pequeno na Internet.

A fim de quantificar o efeito da topologia na resiliência de um rede, Albert e colaboradores (5) simularam falhas aleatórias e ataques em redes geradas pelos modelos livre de escala (SF) e grafos aleatórios (ER), apresentados na Seção 2.5. Os resultados são apresentados na Figura 3.4. Podemos observar novamente que as redes do tipo livre de escala são altamente tolerantes à falhas aleatórias, enquanto que os grafos aleatórios são bastante frágeis. No entanto, quando ataques são simulados, o quadro se inverte, onde observa-se que as redes livre de escala são altamente vulneráveis à ataques. Portanto, redes livre de escala são altamente robustas à falhas aleatórias, mas são frágeis a ataques dirigidos.

#### 3.2.2 Medida para caracterização da robustez

Robustez é definida (4) ser a insensibilidade dos observáveis de uma dada rede perante alterações estruturais ou topológicas, como por exemplo, remoção aleatória de nós. Sendo assim, a robustez de uma rede complexa pode ser quantificada pela análise da variação de sua organização quando submetida à falhas e ataques. No entanto, essa análise não permite definirmos um índice para quantificar o quando uma rede é robusta. Na verdade, para redes aleatórias, é possível quantificar o nível de robustez através do coeficiente crítico  $p_c$ , que representa a fração de vértices removidos a partir do qual a rede se fragmenta. Isto é, para  $p > p_c$ , a rede se quebra em mais de um componente conectado. No entanto, pesquisadores de Israel verificaram em 2001 que tal expoente não existe em redes livre de escala, indicando que tais redes são altamente robustas à falhas aleatórias (84).

Apenas recentemente, foi proposto um índice para quantificar a robustez de uma rede complexas baseado em métodos da mecânica estatística e da teoria ergódica. A ideia é que características macroscópicas da rede são mantidas constantes para um tempo relativamente longo, contudo, isto não implica que microestados sejam estáticos. É a diversidade e incerteza de processos microscópicos que determinam a resiliência de estados macroscópicos estacionários contra perturbações. No contexto da teoria ergódica de sistemas dinâmicos, esta incerteza é quantificada pela entropia dinâmica  $H_d$  (entropia de Kolmogorov-Sinai). De acordo o teorema da flutuação para as redes (4), mudanças na entropia dinâmica são correlacionadas positivamente com mudanças na resiliência do sistema macroscópico contra perturbações microscópicas. Vamos assumir um processo estocástico definido por uma matriz de Markov  $P = (p_{ij})$ . Tal processo descreve a taxa de transição entre quaisquer estados  $i \rightarrow j$  ( $p_{ij} \ge 0 e \sum_j p_{ij} = 1$ ). Desta forma, a entropia dinâmica de um processo de Markov caracteriza a diversidade de caminhos possíveis e, por causa do teorema da flutuação (4), é relacionada à resposta do sistema contra perturbações. A definição da entropia dinâmica para um processo de Markov P, é dada por (85):

$$H_d = -\sum_{ij} \pi_i p_{ij} \log(p_{ij}), \qquad (3.10)$$

onde  $p_{ij}$  foi obtido da matriz de probabilidade de transição *P*, cujos elementos  $p_{i,j} = a_{i,j}/k_j$ . Isto é, cada elemento  $p_{ij}$  representa a probabilidade de alguém que caminha aleatoriamente ir do nó *i* ao nó *j* em um passo. Os termos  $\pi_i$  são as componentes da distribuição estacionária, isto é,  $\pi$  é definido como o autovetor do lado esquerdo associado com o maior autovalor 1 da matriz estocástica *P*,  $\pi P = \pi$ . Esta distribuição estacionária caracteriza o comportamento invariável a longo prazo do processo de Markov que descrevemos por pela matriz *P*. Também podemos classificar os elementos da rede individualmente, pela decomposição da entropia da rede em contribuições de todos os vértices, isto é,

$$H_d(i) = -\pi_i \sum_j P(i, j) \log(P(i, j)),$$
(3.11)

Esta decomposição sugere que, os elementos da rede com um alta contribuição para a entropia global têm um grande efeito na resiliência e funcionalidade da rede, quando removidos.

A Figura 3.5 apresenta topologias associadas a diferentes valores de entropia dinâmica. Nesse caso, uma rede onde todos os vértices são conectados a um nó central possui uma robustez maior, em termos de remoções aleatórias, do que uma rede regular formada por uma sequência de vértices conectados. Isso se deve ao fato de que a probabilidade de se selecionar o vértice central é muito menor do que remover um vértice pouco conectado, enquanto que no caso da rede regular, a remoção de qualquer vértice leva à decomposição da rede em diversos componentes. Sendo assim, a entropia dinâmica mostra que redes livre de escala são mais robustas do que redes aleatórias.





**Figura 3.5** – A entropia dinâmica,  $H_d$ , depende da combinação de várias características da rede, assim como a distribuição do grau e o menor caminho médio. Nesta ilustração, todas as redes possuem o mesmo número de nós (N = 100) e *links* (M = 200). Figura extraída de (4).

Os conceitos de falhas e ataques serão empregados na análise de redes de interação de proteínas e malhas rodoviárias, onde vamos comparar redes de diversas espécies e países. De fato, esse tipo de análise é fundamental para se verificar se organismos mais complexos apresentam maior índice de resiliência e, ao mesmo tempo, verificar como as ligações entre cidades influencia no transporte de pessoas e mercadorias.

# 4 ROBUSTEZ DE REDES DE INTERAÇÃO PROTEÍNA-PROTEÍNA

### 4.1 Redes biológicas

Nos últimos anos, o estudo da biologia molecular tem recebido um ponto de vista interdisciplinar, focado na interação de sistemas biológicos, levando a uma nova perspectiva, isto é, uma visão holística no lugar do reducionismo (86). Processos biológicos dependem da interação entre proteínas, bem como proteínas e outras moléculas, que geram redes de interação (86). Proteínas podem interagir a fim de formar um complexo celular, onde cada proteína carrega outra proteína, ou interage brevemente com uma proteína alvo a fim de modificá-la. Mapas de interação de proteínas podem ser naturalmente representados como redes, onde os nós representam as proteínas e as arestas representam as interações (87). Estas redes de interação proteína-proteína tem sido estudadas por um longo tempo (88), mas só recentemente que conceitos teóricos de grafos foram aplicadas na análise de tais redes (89).

Um importante objeto de investigação para pesquisadores de sistemas biológicos tem sido a relação entre a estrutura da rede biológica e suas respectivas funções que controlam o fluxo de informação e regulação de sinais celulares (90, 91). De fato, tem sido verificado que a estrutura da rede de interação proteína-proteína difere de redes aleatórias no senso dos modelos de Erdös e Rényi (48) (seção 2.5.1), mas apresenta uma estrutura que segue o paradigma de *livre de escala* (72) (ver seção 2.5.3), caracterizado pelo fato que a distribuição dos números de conexões seguem uma lei de potência. Na realidade, a distribuição das conexões das redes de interação de proteínas segue uma lei de potência com um corte (*cutoff*) exponencial (54, 92). Este tipo de topologia é particularmente interessante pois permite a existência de *hubs*, isto é, vértices com alto grau, que são conhecidos por desempenhar um papel importante em redes (20). De fato, *hubs* tem uma intrínseca relação com letalidade, conforme verificado em trabalhos recentes (54, 93).

A relação entre a organização e robustez de redes biológicas pode ser investigada através da remoção de genes. Recentemente, a análise da robustez de redes biológicas foi experimentalmente realizada por estudos em resposta direta a desativação de genes (94) ou interferência no RNA (95). Essas experiências são estratégias fundamentais para a compreensão da organização global das funções genéticas (94).

Computacionalmente, o exame das perturbações pode ser executado por remoção aleatória ou preferencial de proteínas em redes de interação de proteínas (5). No primeiro caso, todas proteínas tem a mesma probabilidade de serem removidas, enquanto no último caso, proteínas são selecionadas de acordo com um dado atributo, tal como o número de conexões. Um prévio trabalho analisou a robustez de proteínas da *Saccharomyces cerevisiae, Caenorhabditis elegans* e *Drosophila melanogaster*, onde os autores descreveram a variação do diâmetro da rede sobre falhas, isto é, remoção aleatória, e ataques, removendo as proteínas mais conectadas seguindo a ordem do números de conexões de cada uma delas (96).

Contudo, este trabalho mostra somente um quadro da resiliência da rede sem prover qualquer explicação sobre que característica topológica que está favorecendo a resiliência. Além disso, somente uma particular propriedade particular de redes, ou seja, o diâmetro, foi analisado. De fato, assumindo que a organização das redes é um produto da seleção natural, é esperado que a estrutura de redes biológicas de espécies diferindo em termos da complexidade deve ser diversificada de alguma forma. Aqui, apresentamos, pela primeira vez, a quantificação de quais propriedades estruturais favorecem a resiliência de redes de interação proteína-proteína.
#### 4.2 Bases de dados

A fim de comparar redes de diferentes espécies, consideramos duas base de dados de diferentes fontes para cada espécie. Esta adoção permite obter uma comparação mais confiável, um vez que as base de dados fornecem redes com diferentes número de proteínas a conexões. Para a levedura Saccharomyces cerevisiae (fig. 4.1 a), nós adotamos a base de dados do Center for Cancer Systems Biology (CCSB), que prove alta qualidade em informação de interação binária (97). Esta base de dados é considerada uma segunda geração de alta qualidade, cobrindo cerca de 20% de todas as interações binárias da levedura (97). Uma das base de dados adotadas cobre 1809 interações entre 1278 proteínas (chamamos esta rede de Sce-CCSB). A segunda base de dados está baseada no experimento de Ito e Uetz e é composta por 2930 interações entre 2018 proteínas (chamamos de Sce-Union) (98). No caso do verme Caenorhabditis elegans (fig. 4.1 b), levamos em conta duas bases de dados: (i) a Worm Interactome (99), do CCSB (chamada Cel-Wi), que é composta por 1496 proteínas e 1816 interações (do Vital Lab 2007), e (ii) a base de dados Biogrid (100), versão 3.0.65 (chamada Cel-BG), composta por 3518 proteínas e 6672 interações. As bases de dados da Drosophila melanogaster (fig. 4.1 c) foram também obtidas da base Biogrid (100), versão 3.0.65 (chamada Dme-BG), formada por 7282 nós e 27671 conexões; e do Drosophila Interaction Database (Droid), que inclui dados de interação de proteína gerados no laboratório Finley usando o sistema híbrido para a levedura LexA, principalmente a partir de HTS (101). Esta última base de dados resultou em uma rede (chamada Dme-FI) composta por 1338 nós e 3161 interações. Enfim, para o Homo sapiens (fig. 4.1 d), levamos em conta base de dados de Biogrid (100), também versão 3.0.65 (chamada Hsa-BG), composta por 9527 proteínas e 36877 interações; e a Human Protein Reference Database versão 9 (102), formada por 9612 proteínas e 39128 interações.

Para todas estas bases de dados, nó levamos em conta somente o maior componente para realizar a análise de resiliência. Tal componente foi extraído usando-se os software Cytoscape (103). A Figura 4.1 apresenta uma amostra de cada espécie analisada.



Figura 4.1 – Redes geradas pelo software *Gephi*: a) *Saccharomyces cerevisiae* (Sce-Union), b) *Caenorhabditis elegans* (Cel-BG), c) *Drosophila melanogaster* (Dme-BG) e d) *Homo sapiens* (Hsa-BG).

### 4.3 Caracterização de redes de interação de proteínas

Redes biológicas, como qualquer outra rede complexa, podem ser caracterizadas pelas suas propriedades estruturais e dinâmicas, conforme descrevemos a seguir.

### 4.3.1 Medidas topológicas

Redes de interação de proteínas são tipos de redes complexas não dirigidas (ver seção 2.3.1), formadas por um conjunto de *N* vértices, as proteínas, conectados por *M* arestas, indicando as interações entre as proteínas. Uma rede pode ser representada por sua matriz adjacência  $\mathscr{A}$ , cujos elementos  $a_{ij}$  e  $a_{ji}$  são iguais a um sempre que houver uma conexão entre as proteínas *i* e *j*, e igual a zero caso contrário. A topologia de redes de interação pode ser descrita por um conjunto de medidas de redes (71).

Para descrever estas redes, retomaremos as medidas estudadas nos capítulos anteriores. A conectividade de uma proteína i, sendo  $k_i$ , é igual ao número de arestas conectadas a esta proteína e pode ser calculado como na equação 2.6. A conectividade média da rede é a média de  $k_i$  para todos os vértices na rede, dada pela equação 2.7, e quantifica a densidade de conexões. A distribuição de grau P(k) dá a probabilidade de se encontrar uma proteína com grau k em uma rede. Redes de interação proteína-proteína apresentam ciclos de ordem três, e uma maneira para caracterizar a presença de laços (loops) desta ordem é através do coeficiente de aglomeração (*clustering coefficient*), onde cc<sub>i</sub> é o coeficiente de aglomeração da proteína i, definido conforme a equação 2.8, e sua respectiva medida global  $\langle cc \rangle$ , definida pela equação 2.9. A média conectividade dos vizinhos, r(i) é descrita na Seção 3.1.2. O menor caminho médio,  $\ell$ , é calculado levando-se em conta a menor distância entre cada par de proteínas em uma rede, assim como o diâmetro d de uma rede, corresponde ao maior dentre todos os menores comprimentos entre qualquer par de proteínas (ver Seção 2.4.4). O coeficiente de assortatividade, que chamamos de s, mede a correlação entre as conectividades da proteínas, e é obtido por calcular o coeficiente de correlação de Pearson entre os graus de cada par de nós. Se s > 0 a rede é assortativa; se s < 0, a rede é disassortativa (ver Seção 3.1.1).

#### 4.3.2 Medidas relacionadas a resiliência

A resiliência de proteínas pode ser quantificadas por medidas relacionadas a entropia (81, 104). A entropia da distribuição de grau tem revelado ser uma eficiente medida de resiliência de redes sob perturbação (81). Esta medida provê uma média da heterogeneidade da rede, uma vez que mede a diversidade da distribuição das ligações (80). A entropia H da distribuição de grau P(k) é definida pela equação 3.9. E para o caso de redes de interação de proteínas, que apresentam uma lei de potência na distribuição de grau com um limite (cutoff) exponencial, como o expoente de escala  $\gamma$  aumenta, a rede torna-se menos heterogênea e, como resultado, um baixo valor de entropia deve ser observado. Outra medida adotada relacionada a resiliência é a entropia dinâmica, que é baseada em dinâmica estocástica de caminhadas aleatórias. Esta particular dinâmica tem sido usada para modelagem de ativações sequenciais de proteínas, tais como transdução de sinal (105), como descrito em (93). A entropia dinâmica de um processo de Markov caracteriza a diversidade de possíveis caminhos, e é relacionada a resposta do sistema a perturbações (85). Esta medida foi usada para caracterização de proteínas nas redes da levedura, revelando que proteínas com larga contribuição para a entropia da rede são preferencialmente letais (104). A entropia dinâmica  $H_d$  é calculada pela equação 3.10. Uma vez que a entropia da distribuição de grau e a entropia dinâmica são medidas bem estabelecida (81, 104), ambas foram usadas na quantificação da resiliência das redes de interação de proteínas analisadas.

#### 4.4 Resultados obtidos

A distribuição de grau das redes de interação proteína-proteína seguem uma lei de potência com um corte (*cutoff*) exponencial, que é devido ao efeito do tamanho finito (55), descrito por:

$$\mathbf{P}(\mathbf{k}) \simeq k^{-\gamma} e^{-k/k_c} \tag{4.1}$$



Figura 4.2 – Distribuição acumulada do grau das redes de interação de proteína-proteína da S. cerevisiae (Sce-CCSB e Sce-Union), C. elegans (Cel-Wi e Cel-BG), D. melanogaster (Dme-BG e Dme-Fl) e H. sapiens (Hsa-BG e Hsa-CCSB).

A Figura 4.2 mostra a distribuição cumulativa ( $P_{cum}(k) = \int_{k}^{\infty} P(x) dx$ ) obtida para todas as oito redes. Os valores de  $\gamma$  (obtido usando o método dos mínimos quadrados) e das outras medidas são mostrados na Tabela 4.1. A média do coeficiente de aglomeração,  $\langle cc \rangle$ , é relativamente pequena para todas as redes, indicando que duas proteínas conectadas tendem a não compartilhar o mesmo conjunto de vizinhos. Além disso, os valores da média do comprimento dos menores caminhos indica que todas redes de proteínas consideradas são small-world, que pode ser uma consequência da presença de hubs. É interessante notar que os menores valores de  $\ell$  tendem a ser obtidos para as espécies mais complexas. De fato, quanto menor é a média do caminho mais curto, mais rápida é a comunicação entre os nós, o que é um aspecto que pode ser relacionado a resiliência (106). O diâmetro das redes parece ser preservado para todas as espécies, variando de 12 a 16. A média do grau dos vizinhos também é maior para as espécies mais complexas, com a H. sapiens e D. melanogaster apresentando os mais altos valores. Todas as redes investigadas são disassortativas exibindo os menores coeficientes s para o S. cerevisiae e C. elegans. Comportamento similar foi observado para o coeficiente da distribuição de grau. Ademais, as redes de interação de proteínas da S. cerevisiae e C. elegans têm os menores expoente de grau  $\gamma$ , que significa que a topologia destas redes são influenciadas principalmente pelas proteínas

Tabela 4.1 – Conjunto de medidas obtidas para a redes das quatros espécies consideradas, S. cerevisiae (Sce-CCSB e Sce-Union), C. elegans (Cel-Wi e Cel-BG), D. melanogaster (Dme-BG e Dme-Fl) e H. sapiens (Hsa-BG e Hsa-CCSB).

Species	$\langle cc \rangle$	l	d	r	S	γ	Н	$H_d$
S. cerevisiae (Sce-CCSB)	0.056	5.35	14	3.08	-0.16	2.47	2.56	2.47
S. cerevisiae (Sce-Union)	0.057	5.61	14	3.06	-0.11	2.47	2.59	2.41
C. elegans (Cel-Wi)	0.017	5.72	16	2.71	-0.25	2.53	2.20	2.37
C. elegans (Cel-BG)	0.060	4.32	13	3.85	-0.17	2.17	2.55	3.50
D. melanogaster (Dme-BG)	0.046	4.32	11	6.93	-0.04	1.77	3.88	3.79
D. melanogaster (Dme-Fl)	0.043	4.50	12	5.02	-0.15	2.05	3.09	3.53
H. sapiens (Hsa-BG)	0.157	4.48	14	6.76	-0.07	1.96	3.79	3.86
H. sapiens (Hsa-CCSB)	0.106	4.21	14	8.00	-0.04	1.91	4.09	4.11

altamente conectadas. A seguir, nós verificamos que esta propriedade parece ser fundamental para definir o nível de resiliência da rede.

#### 4.4.1 Falhas aleatórias

Nós iniciamos a análise dinâmica por investigar efeito da remoção aleatória de proteínas das redes das quatros espécies consideradas. Esta tarefa foi alcançada sem qualquer regra preferencial, isto é, todas as proteínas tiveram a mesma probabilidade de serem selecionadas para remoção em uma rede. A Figure 4.3 apresenta a variação do tamanho da maior componente, diâmetro e menor caminho médio para cada rede. Estas medidas são as mais frequentemente adotadas na literatura para análise de resiliência (20), uma vez que elas quantificam a capacidade da rede de manter o seu comportamento normal sob perturbações internas ou externas. Note que cada medida da rede foi normalizada com respeito ao seu valor original, isto é, a medida calculada na rede sem perturbação. Esta abordagem foi adotada, uma vez que, as redes apresentam diferentes números de vertices e arestas. Além disso, as medidas estudadas dependem do tamanho da rede e da densidade de conexões. Portanto, a normalização nos permitiu investigar somente a variação da medida evitando efeitos impostos pelo tamanho da rede e da densidade de conexão. Na Figura 4.3, isto pode ser observado que todas as redes apresentam



Figura 4.3 – Falhas aleatórias em redes de interação proteína-proteína. A variação: (a) do tamanho da maior componente, (b) do diâmetro e (c) da média do comprimento do menor caminho, como função da fração dos vértices removidos para redes de proteínas da *S. cerevisiae* (Sce-CCSB e Sce-Union), *C. elegans* (Cel-Wi e Cel-BG), *D. melanogaster* (Dme-BG e Dme-Fl) e *H. sapiens* (Hsa-BG and Hsa-CCSB). Cada ponto é uma média sobre 1000 simulações.

comportamento similar. A principal diferença ocorreu para o maior componente da *S. cerevisiae* e *C. elegans*. E também, o diâmetro e a média do menor caminho de todas as redes apresentam variações muito pequenas. Portanto, as redes de todas as espécies exibem similar resiliência

contra falhas aleatórias, não importando a complexidade do organismo. Na verdade, esta é uma consequência do fato que todas estas redes são de *livre-escala* apresentando distribuição de grau que segue uma lei de potência com um limite (*cutoff*) exponencial, como mostrado na Figura 4.2. Este resultados obtidos corroboram com observações experimentais, que mostraram que um grande número de perturbações não resulta em qualquer variação fenotípica sob uma da condição experimental (94, 95).

#### 4.4.2 Ataque intencional

A resiliência das redes de interação de proteínas foi também investigada com respeito a ataques intencionais. Neste caso, os nós mais conectados foram removidos sequencialmente, de acordo com seu grau, e uma respectiva medida foi calculada. Nós realizamos tal análise removendo até 5% do total do número de vértices e obtendo a variação relativa da maior componente, diâmetro da rede e média do comprimento do menor caminho. Como é feito no caso de falha aleatória, a variação relativa foi obtida pela relação entre a medida calculada para uma porcentagem f de nós removidos e a medida original. A Figura 4.4 apresenta o resultado obtido. Diferentemente da falha aleatória, a variação estrutural das redes sob ataques depende da complexidade do organismo. Em todos os casos, quanto mais complexo um organismo, mais resiliente ele tende ser a ataques intencionais. A interação proteína-proteína da *D. melanogaster* e *H. sapiens* são as mais robustas, enquanto a rede da *S. cerevisiae* é a menos resiliente.

#### 4.4.3 Relação entre a estrutura da rede e a resiliência

Medidas relacionadas a entropia, como a estrutural (81) e entropia dinâmica (104), têm sido levadas em conta para quantificação da robustez de rede em redes complexas. Nós adotamos estas medidas a fim de identificar que propriedade das redes contribui mais decisivamente para



Figura 4.4 – Ataques intencionais em redes de interação proteína-proteína. A variação: (a) do tamanho da maior componente, (b) do diâmetro e (c) da média do comprimento do menor caminho, como função da fração dos vértices removidos para redes de proteínas da *S. cerevisiae* (Sce-CCSB e Sce-Union), *C. elegans* (Cel-Wi and Cel-BG), *D. melanogaster* (Dme-BG e Dme-Fl) e *H. sapiens* (Hsa-BG e Hsa-CCSB). Cada ponto é uma média sobre 1000 simulações.

a resiliência das redes de interação de proteínas. A Tabela 4.1 apresenta os valores de entropia obtidos para cada rede. Comparando estes valores com as curvas das Figuras 4.3 e 4.4, nós podemos observar um bom acordo entre eles, isto é, redes mais resilientes tendem a ter maior



**Figura 4.5** – A correlação entre as medidas de entropia e cada medida é indicada por  $\rho$ . Os pontos em pretos são relacionados a entropia dinâmica  $H_d$ , e em cinza a entropia da distribuição de grau H. As medidas estruturais são o coeficiente de aglomeração médio  $\langle cc \rangle$ , a média de comprimento dos menores caminhos  $\ell$ , o diâmetro da rede d, o grau médio dos vizinhos r, o coeficiente de assortatividade s e o coeficiente de escala da lei de potência da distribuição de grau  $\gamma$ .

entropias dinâmica e estrutural. Deste modo, nós consideremos um conjunto de seis medidas e calculamos seus coeficientes de correlação de Pearson com a entropia da distribuição de conectividade e a entropia dinâmica. Esta abordagem permite a quantificar relação entre cada característica estrutural e a robustez global da rede. A Figura 4.5 apresenta a relação obtida entre a entropia dinâmica e a medida adotada. A correlação entre a entropia da distribuição de grau e cada medida ( $\rho$ ), bem como entre a entropia dinâmica e cada medida da rede, também são dadas nesta figura.

Ao comparar os valores de correlação, pode-se observar que o grau médio dos vizinhos, r, e o coeficiente da lei de potência,  $\gamma$ , são as medidas que apresenta as maiores correlações com as entropias. Esta duas propriedades estão potencialmente contribuindo de uma forma particularmente forte para a resiliência. A fim de verificar a diferença entre o grau médio do vizinhos de todas as espécies consideradas, nós calculamos as distribuições destas medidas levando-se em conta somente os hubs removidos, que representam 5% do total do número de proteínas. A Figura 4.6 mostra os os resultados obtidos. Os valores das medianas para cada distribuição também são indicados nesta figura. Adotamos a mediana ao invés da média porque as distribuições não são simétricas (107). Pode-se observar que *H. sapiens* e *D. melanogaster* apresentam hubs com maior grau médio dos vizinhos do que a *S. cerevisiae* e a *C. elegans*. Portanto, nossos resultados sugerem que este mecanismo de reforçar as conexões dos vizinhos dos hubs é particularmente responsável para a maior resiliência dos organismos mais evoluídos. Na verdade, os maiores valores do grau médio dos vizinhos dos hubs indicam a presença de conexões alternativas entre vizinhos de tais nós, o que favorece a resiliência das redes. Como resultado, a alta densidade de conexões entre os nós conectados a hubs tendem a minimizar as remoções dos hubs.

Embora o grau médio dos vizinhos e o coeficiente da lei de potência apresentam a maior correlação com os coeficientes de entropia, as outras medidas também apresentam uma correlação significativa. O comprimento do menor caminho é fortemente correlacionado, negativamente, com a entropia dinâmica, indicando que menores comprimentos dos caminhos mais curtos favorecem a resiliência. De fato, a alta densidade de conexões entre os vizinhos de cada nó tende a criar caminhos alternativos entre pares de nós reduzindo a média comprimento dos menores caminhos da rede. Já o diâmetro apresenta uma pequena correlação com as medidas de entropia. É interessante notar que os coeficientes de assortatividade não são fortemente correlacionados com as medidas de entropia, indicando que esta característica não está contribuindo significativamente para a resiliência. Na verdade, quanto menor for esse coeficiente, menos robusta é a rede. Esse resultado concorda com outros trabalhos, que verificaram que redes assortativas são consideráveis mais robustas contra a remoção de vértices do que redes disassortativas (76).

Em resumo, os resultados apresentados neste capítulo são fundamentais para o entendimento da relação entre a organização de grande escala das redes de interação de proteínaproteína e a respectivas funções biológicas. A regra particularmente investigada, ou seja, a resiliência, está relacionada à evolução biológica e organização do organismo (108). Assim, a identificação de que regra está favorecendo a evolução, pode ajudar o desenvolvimento de modelos computacionais de evolução de rede de proteínas, tais como (109), e entender quais mecanismos são usados durante a evolução natural para produzir arquiteturas robustas.



Figura 4.6 – Distribuição do grau médio dos vizinhos para os hubs removidos da S. cerevisiae (Sce-CCSB e Sce-Union), C. elegans (Cel-Wi and Cel-BG), D. melanogaster (Dme-BG e Dme-Fl) e H. sapiens (Hsa-BG e Hsa-CCSB). A mediana de cada distribuição, m, também é indicada nos respectivos gráficos.

## 5 ROBUSTEZ, ACESSIBILIDADE E CENTRALIDADE DE MALHAS RODOVIÁRIAS

O transporte de pessoas e mercadorias entre diferentes cidades constitui uma atividade fundamental tanto do ponto de vista econômico como social. Cidades são ligadas por estradas e a organização da malha rodoviária tem influência fundamental sobre o tráfego. Na verdade, estradas são construídas de modo a favorecer a conexão de cidades próximas e minimizar os recursos disponíveis. Sendo assim, é natural representar os mapas rodoviários como redes, de modo que as cidades ou os cruzamentos entre rodovias, constituem os vértices, e as estradas são as arestas que conectam tais vértices. Apesar da representação dos vértices como cidades parecer mais natural, ela apresenta uma limitação no estudo de transporte em redes, conforme observamos na Figura 5.1. Desse modo, a representação que leva em conta os cruzamentos permite uma análise mais precisa do fluxo em malhas rodoviárias e, portanto, usaremos tal representação neste estudo das redes de rodovias.

#### 5.1 Aquisição de dados

Desde do final do século passado, o uso do computador tem sido uma ferramenta indispensável na cartografia para a produção de mapas. Em cartografia, quando se trata de aquisição de dados, a maior parte tem sido adquirida por Sistemas de Informação Geográfica ou GIS (*Geographic Information Systems*), sistemas computacionais de captura, armazenagem e análise de



Figura 5.1 – Mapa rodoviário da Índia com: (a) pontos representando as cidades localizadas próximas das rodovias (nem todos pontos são intersecções de vias), (b) pontos representando apenas os cruzamentos.

dados espaciais. Um GIS pode ser utilizado por várias áreas do conhecimento como a geografia, planejamento urbano, sensoriamento remoto (110), e seus dados representam objetos reais como rodovias, terras habitadas, rios e outros detalhes.

Para este estudo foram adquiridos arquivos digitais de mapas geográficos de países no formato de documento portável (*pdf*), disponíveis na Internet (alguns mapas foram obtidos do site das Nações Unidas, http://www.un.org/, ou no caso do Brasil pelo site do Ministério dos Transportes, http://www2.transportes.gov.br/). Posteriormente foram extraídas apenas as malhas rodoviárias destes mapas, devido ao fato de existirem outras representações de objetos que são irrelevantes neste caso. Considerando apenas as principais rodovias para a análise, logo que estas captam o fluxo de transporte das cidades mais importantes (não estamos interessados nas vias secundárias que partem de rodovias principais para se ligarem a pequenas cidades), então é gerado uma imagem em tons de cinza, onde todas as estradas passam a possuir uma mesma cor, e um algoritmo que percorre os pixels desta malha de rodovias é executado. Inicialmente o algoritmo, que percorre a malha de rodovias, identifica todos os pontos de intersecções, através da esqueletização das vias (deixando-as com a largura de um pixel, por pegar apenas a linha central entre as bordas de uma via) e convolução com um filtro que identifica as junções



**Figura 5.2** – (a) Mapa de interligação de cruzamentos do Brasil. (b) Cruzamentos rotulados por cores, através do algoritmo de propagação de rótulos.

de vias (um filtro acerto e erro de 8-conectados capaz de encontrar as junções diagonais e ortogonais, isto ocorre somente quando 4 pixels vizinhos são encontrados), então, estes cruzamentos são marcados com pontos em preto. Tendo identificado as intersecções, o algoritmo faz a rotulação de todos estes pontos de cruzamento e, em seguida, faz a propagação de cada rótulo (por difusão) ao longo das vias. Quando um rótulo *i*, ao ser propagado, encontra-se com outro rótulo *j*, é identificado uma ligação entre estes cruzamentos (vértices) na matriz de adjacência  $\mathscr{A}$ , fazendo  $a_{ij} = 1$ . O final da execução do algoritmo tem como saída uma lista de conexões da rede de cruzamentos, e uma imagem gerada com o mapa da interligação dos cruzamentos já rotulados e diferenciados por cores (ver Figura 5.2).

#### 5.2 Resultados

A fim de analisar a robustez, acessibilidade e centralidade de redes de rodovias, utilizamos bases de sete países diferentes, isto é, Brasil, Portugal, Polônia, Romênia, Austrália, Índia e África do Sul. As características estruturais de tais redes são apresentadas na Tabela 5.1. Notamos que as redes não são do tipo mundo pequeno, uma vez os valores do diâmetro e média amostral dos menores caminho são maiores do que o observados em redes livre de escala (ver Tabela 2.1). De fato, as malhas rodoviárias são redes do tipo geográfica e as conexões são estabelecidas em um espaço dimensional de modo a favorecer a conexão entre cidades próximas, como observado no modelo apresentado na Referência (111). As distribuições de conectividade obtidas nas malhas rodoviárias do sete países analisados são apresentadas na Figura 5.3, onde nota-se que tais distribuições são, aproximadamente, do tipo Poisson (ver Figura 2.8(b)).

Assumindo que o tráfego nas rodovias ocorre segundo caminhadas aleatórias, podemos estudar diversas propriedades dinâmicas de malhas rodoviárias, com robustez e acessibilidade. No primeiro caso, a robustez de uma rede pode ser quantificada através dos conceitos de entropia dinâmica, discutidas na Seção 3.2.2. Nesse caso, a contribuição de cada vértice para a robustez das redes é dado por (4)

$$H_d(i) = -\pi_i \sum_j P(i, j) \log(P(i, j)),$$
(5.1)

onde  $\pi_i$  representa a probabilidade de transição estacionária de uma cadeia de Markov e *P* é a matriz de probabilidade de transição (112). Quanto maior o valor dessa medida, mais um vértice contribui para a resiliência da rede. Mais especificamente, os vértices com maior valor  $H_d(i)$  são aqueles cuja remoção causa maior variação nas propriedades estruturais de uma rede. Os resultados são mostrados na Figura 5.4. No caso do Brasil, os cruzamentos próximos da cidade de Curitiba, que ligam as cidades da região Sul e Sudeste, e São Paulo são os mais importan-

**Tabela 5.1** – Propriedades topológicas das redes utilizadas: número de vértices (*N*), número de conexões (*M*), coeficiente de aglomeração médio ( $\langle cc \rangle$ ), média dos menores caminhos ( $\ell$ ), diâmetro (*d*), média da conectividade média dos vizinhos de cada vértice (*r*).

País	N	М	$\langle cc \rangle$	$\ell$	d	r
Austrália	816	1157	0,05	19,5	46	2,8
Portugal	295	456	0,09	12,4	37	3,1
Polônia	60	83	0,05	4,9	12	2,7
Brasil	800	1215	0,07	17,9	50	3,0
Índia	572	789	0,05	17,8	49	2,8
África do Sul	84	104	0,01	7,1	18	2,5
Romênia	224	370	0,11	8,8	21	3,3



Figura 5.3 – Distribuição das conexões das redes de rodovias analisadas.

tes para a resistência da malha rodoviária. De maneira similar, na Austrália, os cruzamentos próximos de Melbourne e Camberra são os mais importantes para manutenção da estrutura de comunicação da rede. Observe que cidades importantes, como Sydney, na Austrália, e as capitais do Nordeste brasileiro, que estão localizadas no litoral, não são fundamentais para a robustez da rede, embora sejam importantes centros urbanos. De fato, os resultados apresentados na Figura 5.4 sugerem que os vértices que menos contribuem para o aumento da entropia da rede são aqueles localizados na borda das redes, ou seja, nas fronteiras dos países.

Além da análise local, é possível comparar as redes de rodovias globalmente. Nesse caso, utilizamos medida de entropia dinâmica e entropia da distribuição das conexões, definidas anteriormente, isto é, Equações 3.10 e 3.9. Os resultados obtidos são apresentados na Tabela 5.2. Conforme podemos observar, os países com maior valor de entropia dinâmica são Romênia, Portugal e Brasil. Portanto, esses países são os mais resistentes a falhas na malha rodoviária, de modo que o tráfego possa ser mantido por caminhos alternativos.

Ainda considerando caminhadas aleatórias, é possível analisar a acessibilidade em redes de rodovias. A acessibilidade está relacionada à facilidade de encontrar um certo vértice em uma rede. Ou seja, tal medida quantifica o quanto uma cidade é acessível em um país. Uma possível medida para quantificar essa propriedade de uma rede complexa foi definida recentemente de maneira multi-escala (113). Mais especificamente, a medida é calculada de acordo com um parâmetro h, que representa a distância a partir de um vértice onde a medida é calculada. Por





Figura 5.4 – Entropia dinâmica das redes de rodovias estudadas: (a) Austrália, (b) Brasil, (c) Índia, (d) Romênia, (e) Portugal, (f) Polônia e (g) África do Sul. As cores indicam os valores da entropia dinâmica de cada vértice nas redes.

exemplo, no caso h = 2, não levamos em conta os vizinhos imediatos de um vértice, mas sim os vizinhos dos vizinhos para determinar a acessibilidade. Sendo assim, a acessibilidade foi definida originalmente como:

$$Acc_{h}(i) = \exp\left[-\sum_{j} P_{h}(i,j) \log P_{h}(i,j)\right],$$
(5.2)

onde a soma é feita para  $P_h(i, j) \neq 0$ . A matriz  $P_h(j, i)$  representa a probabilidade de ir do vértice *i* ao vértice *j* em *h* passos. Para caminhadas aleatórias tradicionais, essa matriz de probabilidade de transição pode ser facilmente calculada. Seja

$$P(i,j) = \frac{A(i,j)}{\sum_{j} A(i,j)} = \frac{A(i,j)}{k_i}.$$
(5.3)

então, temos

$$P_h = P^h. (5.4)$$

Observe que a definição da acessibilidade só leva em conta as probabilidades obtidas para caminhos de comprimento h, não levando em conta as demais distâncias.

Com o intuito de considerar todos os caminhos possíveis e com isso evitar o parâmetro hna definição da acessibilidade, propomos uma definição alternativa dessa medida utilizando o conceito de matriz exponencial. Nesse caso, seja A uma matrix  $N \times N$  real ou complexa. Então,

$$e^{A} = \sum_{k=0}^{\infty} \frac{1}{k!} A^{k}.$$
 (5.5)

Essa relação é análoga à função exponencial ordinária. Com isso, podemos definir a acessibilidade em termos da *matriz de probabilidade de comunicação*, que definimos por

$$\mathbf{P} = \sum_{k=0}^{\infty} \frac{1}{k!} P^k, \tag{5.6}$$

cujos elementos representam a probabilidade de ir do vértice *i* ao vértice *j* considerando todos os caminhos de todos os comprimentos possíveis, ponderando cada um deles pelo respectivo

**Tabela 5.2** – Valores da entropia dinâmica ( $H_d$ ), entropia de conectividade (H), acessibilidade (Acc) e betweennesscentrality obtidas para as redes de rodovias estudadas.

País	$H_d$	Н	$\langle B \rangle$	Acc
Austrália	1,58	1,91	$1,51 \times 10^{4}$	8.64
Portugal	1,69	1,82	$0,33 \times 10^{4}$	9.91
Polônia	1,63	2,15	$0,02 \times 10^{4}$	8.42
Brasil	1,69	2,08	$1,35 \times 10^{4}$	10.21
Índia	1,57	2,06	$0,96 \times 10^{4}$	8.10
África do Sul	1,37	1,57	$0,05 \times 10^{4}$	6.23
Romênia	1,79	1,96	$0,17 \times 10^{4}$	11.40



Figura 5.5 – Ilustração do conceito de acessibilidade. Os valores da acessibilidade de cada vértice estão indicados na figura.

comprimento. Assim, definimos a acessibilidade da seguinte forma:

$$Acc(i) = \exp\left[-\sum_{j=1}^{N} \mathbf{P}(i,j) \log \mathbf{P}(i,j)\right].$$
(5.7)

Essa medida apresenta um valor alto para vértices cujas localizações na rede permitem que se comuniquem com outros vértices com a mesma probabilidade. Esse conceito é ilustrado na Figura 5.5. Observe que o vértice em preto, cuja vizinhança é simétrica, apresenta o maior valor de acessibilidade. Por outro lado, os vértices na borda da rede são os que apresentam menor valor de acessibilidade. Esse fenômeno é também observado nas malhas rodoviárias, onde as cidades nas fronteiras dos países apresentam menores valores de acessibilidade. Na verdade, a definição original da acessibilidade também atribui os menores valores para os vértices nas bordas da rede (114).

Devido à essa relação entre a posição do vértice na rede e a acessibilidade, tal propriedade dinâmica é diretamente relacionada com a robustez da rede. De fato, os vértices na fronteira das redes são os que recebem menor fluxo e são os menos acessíveis. Sendo assim, tais vértices

**Tabela 5.3** – Valores da correlação entre a entropia dinâmica local  $(H_d(i))$  e a acessibilidade (Acc(i)).

País	ρ
Austrália	0,85
Portugal	0,82
Polônia	0,86
Brasil	0,87
Índia	0,86
África do Sul	0,84
Romênia	0,85

apresentam um valor baixo de acessibilidade e ao mesmo tempo, quando removidos não causam grande variação na estrutura da rede. Por outro lado, os cruzamentos internos, ou seja, os vértices mais acessíveis, são fundamentais para manter o nível de comunicação. Quando removidos, a comunicação deve ser feita por caminhos alternativos, o que diminui a eficiência da propagação de informação na rede. Para verificarmos essa relação entre acessibilidade e resiliência de forma quantitativa, calculamos o coeficiente de correlação entre essa duas quantidades. Os resultados são apresentados na Tabela 5.3. Verificamos que há uma forte relação entre a acessibilidade e a medida de entropia dinâmica. Como medida global de caracterização das redes, calculamos a média amostral da acessibilidade. Os dados são apresentados na Tabela 5.2. Novamente observamos que quanto maior a acessibilidade média, maior o valor da entropia dinâmica da rede. De fato, o coeficiente de correlação entre tais medidas é igual à 0,97.

Outra propriedade topológica importante que podemos analisar em redes de rodovias é a centralidade. Tal característica pode ser calculada por diferentes métricas, como o grau de intermediação (*betweenness centrality*) e a centralidade espectral (7). Nesse trabalho, analisamos o grau de intermediação  $B_u$ , definido pela Equação 3.6. O grau de intermediação tem sido usado para particionar redes em comunidades (115). Para as redes de rodovias consideradas, os resultados são apresentados na Figura 5.6. Interessante observar que os vértices com maior grau de intermediação são localizados nas regiões centrais de cada país. Embora a acessibilidade também seja relacionada com a centralidade, tal medida não é fortemente relacionada com o grau de intermediação, conforme mostrado na Tabela 5.4. Apenas a Polônia e África do Sul apresentam alto valor de correlação. A baixa correlação entre essa duas medidas para

**Tabela 5.4** – Valores da correlação entre o grau de intermediação  $(B_u)$  e a acessibilidade (Acc(i)).

País	ρ
Austrália	0,35
Portugal	0,54
Polônia	0,85
Brasil	0,56
Índia	0,48
África do Sul	0,87
Romênia	0,68



92



Figura 5.6 – Grau de intermediação para as redes de rodovias estudadas: (a) Austrália, (b) Brasil, (c) Índia, (d) Romênia, (e) Portugal, (f) Polônia e (g) África do Sul. As cores indicam os valores da entropia dinâmica de cada vértice nas redes.

os demais países se deve ao fato de que enquanto o grau de intermediação considera apenas os menores caminhos, a acessibilidade leva em conta todas as caminhadas possíveis entre dois vértices, ponderando cada uma delas pelo seu respectivo comprimento.

Utilizando as medidas de robustez, acessibilidade e centralidade, é possível agrupar os vértices de acordo com a semelhança entre esses três atributos. Nesse caso, cada classe de-

fine a importância de um conjunto de vértices considerando esses três atributos. Utilizando o algoritmo *expectation maximization* (116), determinamos três classes de vértices para cada rede. Os gráficos com as cores representando cada classe são apresentados na Figura 5.7. Os vértices em vermelho são aqueles que apresentam maiores valores de acessibilidade, entropia dinâmica e centralidade, enquanto que os azuis possuem valores intermediários dessas quantidades e os vértices em verde exibem os menores valores. Observamos que os vértices mais internos são os da classe vermelha, enquanto que os mais periféricos da classe verde. Logo, há uma clara relação entre o posicionamento de um cruzamento em um país e sua respectivas propriedades de robustez, acessibilidade e centralidade.

A análise que realizamos nesse capítulo tem diversas implicações práticas. As cidades próximas aos cruzamentos que exibem maiores valores de robustez, acessibilidade e centralidade são aquelas em que deve-se estabelecer pontos de distribuição de mercadorias, pois são os locais onde mais facilmente atingíveis quando se navega pela rede. Enquanto, que as cidades próximas aos cruzamentos com menor acessibilidade, robustez e centralidade são aquelas em que deve-se priorizar o investimento em infra-estrutura rodoviária, de forma a torná-las mais acessíveis. Além disso, nossos resultados podem ser utilizados no planejamento de novas rodovias, de modo que deve-se priorizar as rodovias que contribuam significantemente para o aumento dessas três propriedades de uma rede, isto é, robustez, acessibilidade e centralidade.



Figura 5.7 – Classificação dos vértices de acordo com as medidas de robustez, acessibilidade e centralidade para os países estudados: (a) Austrália, (b) Brasil, (c) Índia, (d) Romênia, (e) Portugal, (f) Polônia e (g) África do Sul. As cores indicam os valores da entropia dinâmica de cada vértice nas redes.

## 6 CONCLUSÕES E TRABALHOS FUTUROS

Os sistemas naturais apresentam a capacidade única de resistir a uma ampla gama de condições adversas extremas, como em cataclismos, por outro lado, produtos projetados pelo homem podem ser corrompidos por causa da falha de um simples componente, como exemplo, um pequeno erro na fiação do circuitos do computador podem desativá-lo completamente. Portanto, cientistas têm reconhecido esta alta resiliência de "projetos da natureza", o que leva-os a explorar as estruturas biológicas em detalhes. Ademais, robustez é um conceito importante do ponto de vista biológico, pois é fundamental entender como uma célula sobrevive e desempenha suas funções sob condições extremas. Tal resiliência está escondida em suas complexas redes de regulação genica e de reações metabólicas. Por isso, a relação entre a estrutura e função de redes biológicas constitui uma questão importante em sistemas biológicos. Particularmente, a estrutura de redes de interação proteína-proteína está relacionada a importantes funções celula-res, tais como a resiliência do organismo contra perturbações aleatórias (ex. mutações).

Nessa dissertação, analisamos a robustez de redes complexas naturais e artificiais com a utilização de um conjunto de medidas estáticas e dinâmicas. Estas medidas foram utilizadas na caracterização e análise dinâmica das redes de interação proteína-proteína das espécies *S. cerevisiae*, *C. elegans*, *D. melanogaster*, *H. sapiens*, e das redes de rodovias dos sete países, isto é, Brasil, Portugal, Polônia, Romênia, Austrália, Índia e África do Sul.

Assim sendo, analisamos no Capítulo 4, a correlação entre a organização topológica e a resiliência de redes de proteínas dos quatro organismos: a levedura *Saccharomyces cerevisiae*, o verme *Caenorhabditis elegans*, a mosca *Drosophila melanogaster* e o *Homo sapiens*. Verifi-

camos que enquanto estas redes exibem variações estruturais similares sobre falha, diferenças significativas surgem quando submetidas a ataques intencionais. Observamos que quanto mais complexa é uma espécie, mais robusta é a sua rede. A adoção de medidas relacionadas com a entropia, para quantificar a resiliência de redes, e suas comparações com seis medidas topológicas revelaram que a densidade de *hubs*, quantificadas pelo coeficiente da lei de potência da distribuição de grau, e a média dos graus dos vizinhos contribuem significativamente para a resiliência das redes. A comparação das estrutura de *hubs* removidos indica que a presença de caminhos alternativos entre a proteínas conectadas aos hubs é um artifício desenvolvido pelas espécies para reforçar a capacidade de resiliência. Esta análise realizada ajudou a entender como a resiliência está subjacente às estruturas das redes, e pode ser aplicada ao desenvolvimento de modelos de redes de proteínas. Como trabalho futuro, outras investigações podem ser realizadas levando em conta outras medidas topológicas da rede e outra dinâmica de remoção, como considerar a remoção dos nós pertencentes a uma dada classe de proteínas. Remoção de arestas também pode ser investigada através de uma abordagem semelhante. Outras medidas para a

No estudo de malhas rodoviárias, apresentamos a medida de acessibilidade, que está relacionada à facilidade de encontrar um dado vértice em uma rede. Relacionando a posição do vértice com a medida de acessibilidade, pudemos observar que os vértices na fronteira destas redes são os que recebem menor fluxo, pois apresentam baixos valores de acessibilidade e, com isso, quando removidos não causam grande variação na estrutura da rede. Mas quando são removidos os vértices mais acessíveis, o transporte deve ser feito por caminhos alternativos, o que diminui a eficiência na distribuição de mercadorias na rede.

Estudamos diversas propriedades dinâmicas de malhas rodoviárias, com robustez (através dos conceitos de entropia dinâmica) e acessibilidade. Achamos a relação entre acessibilidade e resiliência de forma quantitativa, calculando o coeficiente de correlação entre essa duas quantidades, e observamos a forte relação entre estas medidas. Outra propriedade topológica importante que analisamos em redes de rodovias foi a centralidade, através do cálculo do grau de intermediação. Então, utilizando as medidas de robustez, acessibilidade e centralidade, foi

possível agrupar os vértices de acordo com a semelhança entre esses três atributos, através de um algoritmo de classificação. Identificamos as cidades próximas aos cruzamentos com menores valores de acessibilidade, robustez e centralidade. Sendo assim, nossos resultados podem ser utilizados no planejamento de novas rodovias, de modo que se devem priorizar as cidades com menor acessibilidade e índice de robustez a fim de a aumentar significantemente essas três propriedades de uma rede, isto é, robustez, acessibilidade e centralidade. 

# REFERÊNCIAS

1 ZACHARY, W. W. An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, v. 33, n. 4, p. 452–473, 1977. ISSN 0091-7710.

2 WATTS, D.; STROGATZ, S. Collective dynamics of small-world networks. *Nature*, Nature Publishing Group, v. 393, n. 6684, p. 440–442, 1998. ISSN 0028-0836.

3 VÁZQUEZ, A.; PASTOR-SATORRAS, R.; VESPIGNANI, A. Large-scale topological and dynamical properties of the Internet. *Physical Review E*, APS, v. 65, n. 6, p. 66130, 2002. Disponível em: <a href="http://dx.doi.org/10.1103/PhysRevE.65.066130">http://dx.doi.org/10.1103/PhysRevE.65.066130</a>. Acesso em: 12 out. 2010.

4 DEMETRIUS, L.; MANKE, T. Robustness and network evolution–an entropic principle. *Physica A*, Elsevier, v. 346, n. 3-4, p. 682–696, 2005. ISSN 0378-4371.

5 ALBERT, R.; JEONG, H.; BARABÁSI, A.-L. Error and attack tolerance of complex networks. *Nature*, v. 406, p. 378–382, 2000. Disponível em: <a href="http://dx.doi.org/10.1038/35019019">http://dx.doi.org/10.1038/35019019</a>>. Acesso em: 12 out. 2010.

6 BARABÁSI, A.-L. *Linked*: how everything is connected to everything else and what it means for business, science, and everyday life. New York: Penguin Group, 2003.

7 NEWMAN, M. *Networks*: an introduction. New York: Oxford University Press, 2010. ISBN 0199206651.

8 BULLMORE, E.; SPORNS, O. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, Nature Publishing Group, v. 10, n. 3, p. 186–198, 2009.

9 SPORNS, O.; CHIALVO, D. R.; KAISER, M.; HILGETAG, C. C. Organization, development and function of complex brain networks. *Trends in Cognitive Sciences*, Elsevier,

100

v. 8, n. 9, p. 418–425, 2004. ISSN 1364-6613.

10 COSTA, L. F.; JR, O. O.; TRAVIESO, G.; RODRIGUES, F. A.; BOAS, P.; ANTIQUEIRA, L.; VIANA, M.; ROCHA, L. da. Analyzing and modeling real-world phenomena with complex networks: a survey of applications. 2007. Disponível em: <a href="http://arxiv.org/abs/0711.3199">http://arxiv.org/abs/0711.3199</a>. Acesso em: 12 out. 2010.

11 BARRAT, A.; BARTHLEMY, M.; VESPIGNANI, A. *Dynamical processes on complex networks*. New York: Cambridge University Press, 2008. ISBN 0521879507.

12 FALOUTSOS, M.; FALOUTSOS, P.; FALOUTSOS, C. On power-law relationships of the Internet topology. *ACM SIGCOMM Computer Communication Review*, v. 29, n. 4, p. 251–262, 1999.

13 BARABÁSI, A.-L.; ALBERT, R. Emergence of scaling in random networks. *Science*, v. 286, n. 5439, p. 509–12, 1999.

14 BARABÁSI, A.-L.; OLTVAI, Z. N. Network biology: understanding the cells functional organization. *Nature*, v. 5, p. 101–113, 2004.

15 CASTELLANO, C.; FORTUNATO, S.; LORETO, V. Statistical physics of social dynamics. *Reviews of Modern Physics*, APS, v. 81, n. 2, p. 591–646, 2009. ISSN 1539-0756.

16 ALBERT, R.; ALBERT, I.; NAKARADO, G. Structural vulnerability of the North American power grid. *Physical Review E*, APS, v. 69, n. 2, p. 25103, 2004. ISSN 1550-2376. Disponível em: <a href="http://dx.doi.org/10.1103/PhysRevE.69.025103">http://dx.doi.org/10.1103/PhysRevE.69.025103</a>. Acesso em: 10 out. 2010.

17 BORGE-HOLTHOEFER, J.; ARENAS, A. Semantic networks: structure and dynamics. *Entropy*, v. 12, n. 5, p. 1264–1302, 2010.

18 BONANNO, G.; CALDARELLI, G.; LILLO, F.; MICCICHè, S.; VANDEWALLE, N.; MANTEGNA, R. N. Networks of equities in financial markets. *The European Physical Journal B*, v. 38, n. 2, p. 363–371, 2004.

19 TSONIS, A. A.; WANG, G.; SWANSON, K. L.; RODRIGUES, F. A.; COSTA, L. F. Community structure and dynamics in climate networks. *Climate Dynamics*, p. 1–8, 2010. ISSN 0930-7575. Disponível em: <a href="http://dx.doi.org/10.1007/s00382-010-0874-3">http://dx.doi.org/10.1007/s00382-010-0874-3</a>. Acesso em: 12 out. 2010.

20 BOCCALETTI, S.; LATORA, V.; MORENO, Y.; CHAVEZ, M.; HWANG, D. Complex networks: structure and dynamics. *Physics Reports*, Elsevier, v. 424, n. 4-5, p. 175–308, 2006.

21 AMARAL, L.; OTTINO, J. Complex networks: augmenting the framework for the study of complex systems. *The European Physical Journal B*, v. 38, n. 2, p. 147–162, 2004. ISSN 1434-6028.

22 BAR-YAM, Y. *Dynamics of complex systems*. Cambridge: Perseus Books, 2003. ISBN 0201557487.

23 COSTA, L. F.; KAISER, M.; HILGETAG, C. C. Predicting the connectivity of primate cortical networks from topological and spatial node properties. *BMC Systems Biology*, BioMed Central Ltd, v. 1, n. 1, p. 16, 2007. ISSN 1752-0509.

24 WASSERMAN, S.; FAUST, K. *Social network analysis*: methods and applications. Cambridge: Cambridge University Press, 1994. ISBN 0521387078.

25 MCCANN, K. S.; HASTINGS, A.; HUXEL, G. R. Weak trophic interactions and the balance of nature. *Nature*, v. 395, p. 794–798, 1998.

26 GUIMERÀ, R.; AMARAL, L. A. N. Modeling the world-wide airport network. *The European Physical Journal B*, v. 38, n. 2, p. 381–385, 2004. ISSN 1434-6028.

27 GUIMERÀ, R.; MOSSA, S.; TURTSCHI, A.; AMARAL; N., L. A. The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proceedings of the National Academy of Sciences of the United States of America*, v. 102, n. 22, p. 7794–7799, 2005.

28 GASTNER, M. T.; NEWMAN, M. E. J. The spatial structure of networks. *The European Physical Journal B*, v. 49, n. 2, p. 247–252, 2006.

29 FALOUTSOS, M.; FALOUTSOS, P.; FALOUTSOS, C. On power-law relationships of the Internet topology. *Computer Communication Review*, v. 29, n. 4, p. 251–262, 1999.

30 YOOK, S. H.; JEONG, H.; BARABÁSI, A.-L. Modeling the Internet's large-scale topology. *Proceedings of the National Academy of Sciences of the United States of America*, v. 99, p. 13382–13386, 2002.

31 ALBERT, R.; JEONG, H.; BARABÁSI, A. Internet: diameter of the World Wide Web. *Nature*, Nature Publishing Group, v. 401, n. 6749, p. 130–131, 1999. ISSN 0028-0836.

32 BARABÁSI, A.; ALBERT, R.; JEONG, H. Scale-free characteristics of random networks: the topology of the World Wide Web. *Physica A*, Elsevier, v. 281, n. 1-4, p. 69–77, 2000. ISSN 0378-4371.

33 ADAMIC, L. A.; HUBERMAN, B. A. Power-law distribution of the World Wide Web. *Science*, v. 287, n. 5461, p. 2115, 2000. Disponível em: <a href="http://dx.doi.org/10.1126/science.287.5461.2115a">http://dx.doi.org/10.1126/science.287.5461.2115a</a>. Acesso em: 12 out. 2010.

34 COSTA, L. F. What's in a name? *International Journal of Modern Physics C*, v. 15, n. 1, p. 371–379, 2004.

35 ANTIQUEIRA, L.; NUNES, M. G. V.; JR, O. N. O.; COSTA, L. F. Strong correlations between text quality and complex networks features. *Physica A*, v. 373, p. 811–820, 2007.

36 BARABÁSI, A.-L.; JEONG, H.; RAVASZ, R.; NÉDA, Z.; VICSEK, T.; SCHUBERT, A. On the topology of the scientific collaboration networks. *Physica A*, v. 311, p. 590–614, 2002.

37 NEWMAN, M. E. J. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America*, v. 98, n. 2, p. 404–409, 2001.

38 STROGATZ, S. H. Exploring complex networks. *Nature*, v. 410, p. 268–276, 2001.

39 ALBERT, R.; BARABÁSI, A.-L. Statistical mechanics of complex networks. *Reviews of Modern Physics*, v. 74, p. 48–98, 2002.

40 NEWMAN, M. E. J. Structure and function of complex networks. *SIAM Review*, v. 45, n. 2, p. 167–256, 2003.

41 WATTS, D. J. *Small worlds*: the dynamics of networks between order and randomness. Princeton: Princeton University Press, 2003.

42 BORNHOLDT, S.; SCHUSTER, H.; WILEY, J. *Handbook of graphs and networks*: from the genome to the internet. Wheinheim: Wiley-VCH, 2003. ISBN 3527403361.

43 CALDARELLI, G. *Scale-free networks*: complex webs in nature and technology. New York: Oxford University Press, 2007.

44 EULER, L. Solutio problematis ad geometriam situs pertinentis. *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, v. 8, p. 128–140, 1736.

45 WEST, D. B. Introduction to Graph Theory. Upper Saddle River: Prentice Hall, 2001.

46 ERDÖS, P.; RÉNYI, A. On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Science*, v. 5, p. 17–61, 1960.

47 SOLOMONOFF, R.; RAPOPORT, A. Connectivity of random nets. *Bulletin of Mathematical Biology*, Springer, v. 13, n. 2, p. 107–117, 1951. ISSN 0092-8240.

48 ERDÖS, P.; RÉNYI, A. On random graphs I. *Publicationes Mathematicae Debrecen*, v. 6, p. 290–297, 1959.

49 WATTS, D. J. *Six degrees*: the science of a connected age. New York: W.W. Norton & Company, 2004. ISBN 0393325423.

50 MAY, A.; MILNE, D. Effects of alternative road pricing systems on network performance. *Transportation Research Part A*, Elsevier, v. 34, n. 6, p. 407–436, 2000. ISSN 0965-8564.

51 TANENBAUM, A. S. *Computer Networks*. Upper Saddle River: Prentice Hall, 2003. ISBN 0130661023.

52 ROSS, S. M. *Introduction to probability models*. New York: Academic Press, 2007. ISBN 0125980620.

53 JEONG, H.; TOMBOR, B.; ALBERT, R.; OLTVAI, Z.; BARABÁSI, A.-L. The large-scale organization of metabolic networks. *Nature*, Nature Publishing Group, v. 407, n. 6804, p. 651–654, 2000.

54 JEONG, H.; MASON, S.; BARABÁSI, A.-L.; OLTVAI, Z. Lethality and centrality in protein networks. *Nature*, Nature Publishing Group, v. 411, n. 6833, p. 41–42, 2001.

55 NEWMAN, M. Power laws, Pareto distributions and Zipf's law. *Contemporary physics*, Taylor & Francis, v. 46, n. 5, p. 323–351, 2005.

56 PARETO, V. *The mind and society*: trattato di sociologia generale. New York: Harcourt Brace and Company, 1935.

57 PRICE, D. Networks of scientific papers. Science, AAAS, v. 149, n. 3683, p. 510, 1965.

58 AMARAL, L. A. N.; SCALA, A.; BARTHELEMY, M.; STANLEY, H. E. Classes of small-world networks. *Proceedings of the National Academy of Sciences of the United States of America*, v. 97, n. 21, p. 11149–11152, 2000.

59 AIELLO, W.; CHUNG, F. R. K.; LU, L. Random evolution in massive graphs. *IEEE Symposium on Foundations of Computer Science*, p. 510–519, 2001.

60 EBEL, H.; MIELSCH, L. I.; BORNHOLDT, S. Scale-free topology of e-mail networks. *Physical Review E*, v. 66, n. 3, p. 35103, 2002.

61 LILJEROS, F.; EDLING, C.; AMARAL, L. A. N. Sexual networks: implications for the transmission of sexually transmitted infections. *Microbes and Infection*, v. 5, n. 2, p. 189–196, 2003.

62 LILJEROS, F.; EDLING, C. R.; AMARAL, L. A. N.; STANLEY, H. E.; AABERG, Y. The web of human sexual contacts. *Nature*, v. 411, n. 6840, p. 907–908, 2001.

63 BRODER, A.; KUMAR, R.; MAGHOUL, F.; RAGHAVAN, P.; RAJAGOPALAN, S.; STATA, R.; TOMKINS, A.; WIENER, J. Graph structure in the Web. *Computer Networks*, v. 33, n. 1–6, p. 309–320, 2000.

64 REDNER, S. How popular is your paper? an empirical study of the citation distribution. *The European Physical Journal B*, v. 4, n. 2, p. 131–134, 1998.

65 DOROGOVTSEV, S.; MENDES, J. Language as an evolving word web. *Proceedings of the Royal Society of London. Series B:* Biological Sciences, The Royal Society, v. 268, n. 1485, p. 2603, 2001. ISSN 0962-8452.

66 CANCHO, R. F.; SOLé, R. V. The small world of human language. *Proceedings of Royal Society of London B*, v. 268, p. 2261–2266, 2001.

67 CHEN, Q.; CHANG, H.; GOVINDAN, R.; JAMIN, S. The origin of power laws in Internet topologies revisited. In: INFOCOM 2002 ANNUAL JOINT CONFERENCE OF THE IEEE COMPUTER AND COMMUNICATIONS SOCIETIES, 21., 2002. Proceedings... Disponível em: <a href="http://dx.doi.org/10.1109/INFCOM.2002.1019306">http://dx.doi.org/10.1109/INFCOM.2002.1019306</a>>. Acesso em: 12 out. 2010.

68 NEWMAN, M. E. J. Mixing patterns in networks. *Physical Review E*, APS, v. 67, n. 2, p. 26126, 2003.

69 CANCHO, R. F.; JANSSEN, C.; SOLÉ, R. V. Topology of technology graphs: small world patterns in electronic circuits. *Physical Review E*, v. 64, n. 4, p. 46119, 2001.

70 NEWMAN, M. E. J.; PARK, J. Why social networks are different from other types of networks. *Physical Review E*, v. 68, n. 3, 2003.

71 COSTA, L. F.; RODRIGUES, F.; TRAVIESO, G.; BOAS, P. Characterization of complex networks: a survey of measurements. *Advances in Physics*, Taylor & Francis, v. 56, n. 1, p. 167–242, 2007.

72 ALBERT, R.; BARABÁSI, A.-L. Statistical mechanics of complex networks. *Reviews of modern physics*, APS, v. 74, n. 1, p. 47–97, 2002.

73 GROSSMAN, J.; ION, P. On a portion of the well-known collaboration graph. *Congressus Numerantium*, Citeseer, p. 129–132, 1995. ISSN 0384-9864.

74 DAVIS, G.; YOO, M.; BAKER, W. The small world of the American corporate elite, 1982-2001. *Strategic Organization*, Sage Publications, v. 1, n. 3, p. 301, 2003. ISSN 1476-1270.

75 MARTINEZ, N. Artifacts or attributes? Effects of resolution on the Little Rock Lake food web. *Ecological Monographs*, JSTOR, v. 61, n. 4, p. 367–392, 1991. ISSN 0012-9615.

NEWMAN, M. Assortative mixing in networks. *Physical Review Letters*, APS, v. 89,
n. 20, p. 208701, 2002. Disponível em: <a href="http://dx.doi.org/10.1103/PhysRevLett.89.208701">http://dx.doi.org/10.1103/PhysRevLett.89.208701</a>.
Acesso em: 12 out. 2010.

77 FREEMAN, L. C. A set of measures of centrality based on betweenness. *Sociometry*, JSTOR, v. 40, n. 1, p. 35–41, 1977. ISSN 0038-0431.

78 PETZ, D. Entropy, von Neumann and the von Neumann entropy. 2001. Disponível em: <a href="http://arxiv.org/abs/math-ph/0102013">http://arxiv.org/abs/math-ph/0102013</a>>. Acesso em: 12 out. 2010.

79 SHANNON, C. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, ACM, v. 5, n. 1, p. 3–55, 2001. ISSN 1559-1662.

80 SOLÉ, R.; VALVERDE, S. Information theory of complex networks: on evolution and architectural constraints. *Lecture Notes in Physics*, Springer, v. 650, p. 189–207, 2004. ISSN 0075-8450.

81 WANG, B.; TANG, H.; GUO, C.; XIU, Z. Entropy optimization of scale-free networks' robustness to random failures. *Physica A*, Elsevier, v. 363, n. 2, p. 591–596, 2006.

82 CALLAWAY, D.; NEWMAN, M.; STROGATZ, S.; WATTS, D. Network robustness and fragility: percolation on random graphs. *Physical Review Letters*, APS, v. 85, n. 25, p. 5468–5471, 2000. ISSN 1079-7114.
83 PASTOR-SATORRAS, R.; VESPIGNANI, A. Epidemic dynamics and endemic states in complex networks. *Physical Review E*, APS, v. 63, n. 6, p. 66117, 2001. ISSN 1550-2376.

84 COHEN, R.; EREZ, K.; BEN-AVRAHAM, D.; HAVLIN, S. Breakdown of the internet under intentional attack. *Physical Review Letters*, APS, v. 86, n. 16, p. 3682–3685, 2001. ISSN 1079-7114.

85 BILLINGSLEY, P. Ergodic theory and information. New York: Wiley, 1965.

86 BARABÁSI, A.-L.; OLTVAI, Z. Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, Nature Publishing Group, v. 5, n. 2, p. 101–113, 2004.

87 SCHWIKOWSKI, B.; UETZ, P.; FIELDS, S. A network of protein–protein interactions in yeast. *Nature Biotechnology*, Nature Publishing Group, v. 18, n. 12, p. 1257–1261, 2000.

88 MARCOTTE, E.; PELLEGRINI, M.; NG, H.; RICE, D.; YEATES, T.; EISENBERG, D. Detecting protein function and protein-protein interactions from genome sequences. *Science*, AAAS, v. 285, n. 5428, p. 751, 1999.

89 COSTA, L. F.; RODRIGUES, F.; CRISTINO, A. Complex networks: the key to systems biology. *Genetics and Molecular Biology*, SciELO Brasil, v. 31, p. 591–601, 2008.

90 VAZQUEZ, A.; FLAMMINI, A.; MARITAN, A.; VESPIGNANI, A. Global protein function prediction from protein-protein interaction networks. *Nature Biotechnology*, Nature Publishing Group, v. 21, n. 6, p. 697–700, 2003.

91 SHARAN, R.; ULITSKY, I.; SHAMIR, R. Network-based prediction of protein function. *Molecular Systems Biology*, Nature Publishing Group, v. 3, n. 1, 2007. Disponível em: <a href="http://dx.doi.org/10.1038/msb4100129">http://dx.doi.org/10.1038/msb4100129</a>>. Acesso em: 12 out. 2010.

92 COSTA, L. F.; RODRIGUES, F.; TRAVIESO, G. Protein domain connectivity and essentiality. *Applied Physics Letters*, v. 89, p. 174101, 2006. Disponível em: <a href="http://dx.doi.org/10.1063/1.2363142">http://dx.doi.org/10.1063/1.2363142</a> Acesso em: 12 out. 2010.

93 RODRIGUES, F.; COSTA, L. F. Protein lethality investigated in terms of long range dynamical interactions. *Molecular BioSystems*, Royal Society of Chemistry, v. 5, n. 4, p. 385–390, 2009. ISSN 1742-206X.

94 GIAEVER, G. et al. Functional profiling of the Saccharomyces cerevisiae genome. *Nature*, Nature Publishing Group, v. 418, n. 6896, p. 387–391, 2002.

95 KAMATH, R.; FRASER, A.; DONG, Y.; POULIN, G.; DURBIN, R.; GOTTA, M.; KANAPIN, A.; BOT, N. L.; MORENO, S.; SOHRMANN, M.; WELCHMAN, D. P.; ZIPPERLEN, P.; AHRINGER, J. Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. *Nature*, Nature Publishing Group, v. 421, n. 6920, p. 231–237, 2003.

96 LI, D.; LI, J.; OUYANG, S.; WANG, J.; WU, S.; WAN, P.; ZHU, Y.; XU, X.; HE, F. Protein interaction networks of Saccharomyces cerevisiae, Caenorhabditis elegans and Drosophila melanogaster: large-scale organization and robustness. *Proteomics*, John Wiley & Sons, v. 6, n. 2, p. 456–461, 2006.

97 YU, H. et al. High-quality binary protein interaction map of the yeast interactome network. *Science*, AAAS, v. 322, n. 5898, p. 104, 2008.

98 ITO, T.; CHIBA, T.; OZAWA, R.; YOSHIDA, M.; HATTORI, M.; SAKAKI, Y. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proceedings of the National Academy of Sciences of the United States of America*, National Acad Sciences, v. 98, n. 8, p. 4569, 2001.

99 LI, S. et al. A map of the interactome network of the metazoan C. elegans. *Science*, AAAS, v. 303, n. 5657, p. 540–543, 2004.

100 STARK, C.; BREITKREUTZ, B.; REGULY, T.; BOUCHER, L.; BREITKREUTZ, A.; TYERS, M. BioGRID: a general repository for interaction datasets. *Nucleic acids research*, Oxford Univ Press, v. 34, n. 1, p. D535, 2006. ISSN 0305-1048. Disponível em: <a href="http://dx.doi.org/10.1093/nar/gkj109">http://dx.doi.org/10.1093/nar/gkj109</a>>. Acesso em: 12 out. 2010.

101 YU, J.; PACIFICO, S.; LIU, G.; FINLEY, R. DroID: the Drosophila Interactions Database, a comprehensive resource for annotated gene and protein interactions. *BMC genomics*, BioMed Central Ltd, v. 9, n. 1, p. 461, 2008. ISSN 1471-2164. Disponível em: <a href="http://dx.doi.org/10.1186/1471-2164-9-461">http://dx.doi.org/10.1186/1471-2164-9-461</a>). Acesso em: 12 out. 2010.

102 MISHRA, G. R. et al. Human protein reference database-2006 update. *Nucleic acids research*, Oxford Univ Press, v. 34, n. 1, p. D411, 2006. ISSN 0305-1048. Disponível em: <a href="http://dx.doi.org/10.1093/nar/gkj141">http://dx.doi.org/10.1093/nar/gkj141</a>. Acesso em: 12 out. 2010.

103 SHANNON, P.; MARKIEL, A.; OZIER, O.; BALIGA, N.; WANG, J.; RAMAGE, D.; AMIN, N.; SCHWIKOWSKI, B.; IDEKER, T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, Cold Spring Harbor Lab, v. 13, n. 11, p. 2498, 2003. ISSN 1088-9051.

104 MANKE, T.; DEMETRIUS, L.; VINGRON, M. An entropic characterization of protein interaction networks and cellular robustness. *Journal of the Royal Society Interface*, The Royal Society, v. 3, n. 11, p. 843, 2006.

105 SHAYWITZ, A.; DOVE, S.; GREENBERG, M.; HOCHSCHILD, A. Analysis of phosphorylation-dependent protein-protein interactions using a bacterial two-hybrid system. *Science's STKE*, AAAS, v. 2002, n. 142, 2002. Disponível em: <a href="http://dx.doi.org/10.1126/stke.2002.142.pl11">http://dx.doi.org/10.1126/stke.2002.142.pl11</a>. Acesso em: 12 out. 2010.

106 RIVES, A.; GALITSKI, T. Modular organization of cellular networks. *Proceedings of the National Academy of Sciences of the United States of America*, National Acad Sciences, v. 100, n. 3, p. 1128, 2003.

107 BOX, G.; HUNTER, S.; HUNTER, W. *Statistics for experimenters*: design, innovation, and discovery. New York: Wiley, 2005.

108 KITANO, H. Biological robustness. *Nature Reviews Genetics*, Nature Publishing Group, v. 5, n. 11, p. 826–837, 2004.

109 SOLÉ, R.; PASTOR-SATORRAS, R.; SMITH, E.; KEPLER, T. A model of large-scale proteome evolution. *Advances in Complex Systems*, World Scientific Publishing Co. Pte. Ltd., v. 5, n. 1, p. 43–54, 2002.

110 BARBIERI, A. L.; ARRUDA, G. de; RODRIGUES, F. A.; BRUNO, O. M.; COSTA, L. F. An entropy-based approach to automatic image segmentation of satellite images. *Physica A*, v. 390, n. 3, p. 512 – 518, 2011. ISSN 0378-4371.

111 BOAS, P. R. V.; RODRIGUES, F. A.; COSTA, L. F. Modeling worldwide highway networks. *Physics Letters A*, Elsevier, v. 374, n. 1, p. 22–27, 2009. ISSN 0375-9601.

112 TAYLOR, H.; KARLIN, S. *An introduction to stochastic modeling*. San Diego: Academic Press, 1998. ISBN 0126848874.

113 TRAVENÇOLO, B.; COSTA, L. F. Accessibility in complex networks. *Physics Letters A*, Elsevier, v. 373, n. 1, p. 89–95, 2008. ISSN 0375-9601.

114 TRAVENÇOLO, B.; VIANA, M.; COSTA, L. F. Border detection in complex networks. *New Journal of Physics*, IOP Publishing, v. 11, p. 063019, 2009. Disponível em: <a href="http://dx.doi.org/10.1088/1367-2630/11/6/063019">http://dx.doi.org/10.1088/1367-2630/11/6/063019</a>>. Acesso em: 12 out. 2010.

115 GIRVAN, M.; NEWMAN, M. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, National Acad Sciences, v. 99, n. 12, p. 7821, 2002.

116 THEODORIDIS, S.; KOUTROUMBAS, K. *Pattern recognition*. 2nd. ed. New York: Academic Press, 2003.