

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS
DEPARTAMENTO DE FÍSICA E INFORMÁTICA

**Rede Complexa e Criticalidade Auto-Organizada:
Modelos e Aplicações**

Paulo Alexandre de Castro

*Tese apresentada ao Instituto de Física de
São Carlos, Universidade de São Paulo, para
obtenção do Título de Doutor em Ciências:
Física Básica*

ORIENTADOR: *Prof. Dr. Roberto Nicolau Onody*

SÃO CARLOS

2007

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

de Castro, Paulo Alexandre

Rede Complexa e Criticalidade Auto-Organizada: Modelos e Aplicações /
Paulo Alexandre de Castro. — São Carlos, 2007

156 p.

Tese (Doutorado) — Instituto de Física de São Carlos, 2007.

Orientador: Prof. Dr. Roberto Nicolau Onody

1. Criticalidade Auto-Organizada 2. Redes Complexas 3. Rede Barabási-Albert
I. Título.

Aos meus pais, pois me deram a vida. E a minha esposa e amigos, por me dar razão e alegria para vivê-la.

Agradecimentos

Aqui começa uma das partes mais difíceis, mas muito prazerosa desse trabalho, que é lembrar de todos os que me ajudaram nesses últimos anos...

- Primeiramente a Deus, por ter me permitido chegar até aqui.
- À minha querida e amada Marli.
- Ao prestativo e sempre disposto a trabalhar Prof. Onody pela orientação, paciência, dedicação e amizade nesses últimos 6 anos que estivemos trabalhando juntos.
- Ao meu maninho e amigo de todas as horas Perseu e ao meu grande amigo e conselheiro psor Evangelista.
- A todos os colegas de trabalho, e agradecimentos especiais se fazem necessário para a Ana Tereza e Nazareno, Reginaldo, João Vitor, Matheus Lobo, Matheus Lazo, Helena, Cláudia Bonardi, Henrique, Alexandre Colato, Leonardo Maia, Alexandre Rossas, Petrus, Antonio Carlos, Antonio Francisco, Guilherme, Adriano Sousa e Maria Cristina (secretária do grupo da física teórica) pelo companheirismo sempre constante e pelos incontáveis auxílios que me prestaram.
- Às funcionárias da bibliotecas e da secretaria da pós-graduação pela presteza e gentileza no atendimento em todos esses anos.
- Quero agradecer a Wladerez, secretária da pós-graduação, que sempre durante todo o meu doutorado foi muito gentil, paciente, um verdadeiro anjo.
- Aos desenvolvedores do software de processamento de redes Pajek.
- Para concluir, eu gostaria de agradecer a todos que direta ou indiretamente contribuíram para que eu me sentisse tão bem longe de casa.

Este trabalho foi financiado pelo Conselho Nacional de
Desenvolvimento Científico e Tecnológico (CNPq).

Resumo

Modelos e teorias científicas surgem da necessidade do homem entender melhor o funcionamento do mundo em que vive. Constantemente, novos modelos e técnicas são criados com esse objetivo. Uma dessas teorias recentemente desenvolvida é a da Criticalidade Auto-Organizada. No Capítulo 2 desta tese, apresentamos uma breve introdução a Criticalidade Auto-Organizada. Tendo a criticalidade auto-organizada como pano de fundo, no Capítulo 3, estudamos a dinâmica Bak-Sneppen (e diversas variantes) e a comparamos com alguns algoritmos de otimização. Apresentamos no Capítulo 4, uma revisão histórica e conceitual das redes complexas. Revisamos alguns importantes modelos tais como: Erdős-Rényi, Watts-Strogatz, de configuração e Barabási-Albert. No Capítulo 5, estudamos o modelo Barabási-Albert não-linear. Para este modelo, obtivemos uma expressão analítica para a distribuição de conectividades $P(k)$, válida para amplo espectro do espaço de parâmetros. Propusemos também uma forma analítica para o coeficiente de agrupamento, que foi corroborada por nossas simulações numéricas. Verificamos que a rede Barabási-Albert não-linear pode ser assortativa ou desassortativa e que, somente no caso da rede Barabási-Albert linear, ela é não assortativa.

No Capítulo 6, utilizando dados coletados do CD-ROM da revista Placar, construímos uma rede bastante peculiar — a rede do futebol brasileiro. Primeiramente analisamos a rede bipartida formada por jogadores e clubes. Verificamos que a probabilidade de que um jogador tenha participado de M partidas decai exponencialmente com M , ao passo que a probabilidade de que um jogador tenha marcado G gols segue uma lei de potência.

A partir da rede bipartida, construímos a rede unipartida de jogadores, que batizamos de rede de jogadores do futebol brasileiro. Nessa rede, determinamos várias grandezas: o comprimento médio do menor caminho e os coeficientes de agrupamento e de assortatividade. A rede de jogadores de futebol brasileiro nos permitiu analisar a evolução temporal dessas grandezas, uma oportunidade rara em se tratando de redes reais.

Abstract

Models and scientific theories arise from the necessity of the human being to better understand how the world works. Driven by this purpose new models and techniques have been created. For instance, one of these theories recently developed is the Self-Organized Criticality, which is shortly introduced in the Chapter 2 of this thesis. In the framework of the Self-Organized Criticality theory, we investigate the standard Bak-Sneppen dynamics as well some variants of it and compare them with optimization algorithms (Chapter 3). We present a historical and conceptual review of complex networks in the Chapter 4. Some important models like: Erdős-Rényi, Watts-Strogatz, configuration model and Barabási-Albert are revised. In the Chapter 5, we analyze the nonlinear Barabási-Albert model. For this model, we got an analytical expression for the connectivity distribution $P(k)$, which is valid for a wide range of the space parameters. We also proposed an exact analytical expression for the clustering coefficient which corroborates very well with our numerical simulations. The nonlinear Barabási-Albert network can be assortative or disassortative and only in the particular case of the linear Barabási-Albert model, the network is no assortative.

In the Chapter 6, we used collected data from a CD-ROM released by the magazine Placar and constructed a very peculiar network — the Brazilian soccer network. First, we analyzed the bipartite network formed by players and clubs. We find out that the probability of a footballer has played M matches decays exponentially with M , whereas the probability of a footballer to score G gols follows a power-law.

From the bipartite network, we built the unipartite Brazilian soccer players network. For this network, we determined several important quantities: the average shortest path length, the clustering coefficient and the assortative coefficient. We were also able to analyse the time evolution of these quantities — which represents a very rare opportunity in the study of real networks.

Sumário

| | |
|--|-------|
| Agradecimentos | iii |
| Resumo | v |
| Abstract | vi |
| Lista de Figuras | x |
| Lista de Tabelas | xviii |
| Capítulo 1. Introdução | 1 |
| Idéias simples para coisas complicadas | 1 |
| Capítulo 2. Breve Introdução a Criticalidade Auto-Organizada | 5 |
| 2.1 O Todo é Maior do que a Soma das Partes | 5 |
| 2.2 Criticalidade Auto-Organizada | 6 |
| 2.3 Modelo da Pilha de Areia | 7 |
| Capítulo 3. Criticalidade Auto-Organizada, Otimização e Biodiversidade | 12 |
| 3.1 Sistemas Complexos | 12 |
| 3.2 Dinâmicas de Otimização | 14 |
| 3.2.1 Modelo Co-Evolutivo de Bak-Sneppen | 15 |
| 3.2.2 Otimização Extrema: simples e generalizada | 18 |
| 3.3 Dinâmicas Bak-Sneppen e Otimização Extrema Generalizada | 23 |

| | | |
|--|---|----|
| 3.3.1 | Comparação entre as Dinâmicas BS , BSD e GEO-b | 26 |
| 3.3.2 | Dinâmicas BSD e BSR | 30 |
| 3.4 | Resultados | 32 |
| Capítulo 4. Redes, O Que São E Como Surgiram | | 34 |
| 4.1 | Um pouco da vida de Euler e as Pontes de <i>Konigsberg</i> | 34 |
| 4.2 | Nascimento da Ciência das Redes | 38 |
| 4.3 | O Mundo é Pequeno | 41 |
| 4.4 | Ressurgimento da Ciência das Redes | 44 |
| 4.5 | Redes: definição | 45 |
| 4.6 | Tipos de Redes | 47 |
| 4.7 | Propriedades das Redes | 50 |
| 4.7.1 | Distribuição de Conectividades | 50 |
| 4.7.2 | Caminho Geodésico e Diâmetro | 51 |
| 4.7.3 | Efeito de Mundo Pequeno | 52 |
| 4.7.4 | Matriz Adjacência e o Espectro de Autovalores | 54 |
| 4.7.5 | Coefficiente de Agrupamento ou Transitividade | 56 |
| 4.7.6 | Correlação e Conectividade Média dos Primeiros Vizinhos | 58 |
| 4.8 | Erdős e Rényi e os Grafos Aleatórios | 62 |
| 4.9 | Modelo de Configuração | 67 |
| 4.10 | Probabilidade Função Geratriz | 72 |
| 4.10.1 | Propriedades da Função Geratriz | 73 |
| 4.11 | Propriedades de redes não-direcionadas | 77 |
| 4.12 | Watts e Strogatz e as Redes Mundo Pequeno | 83 |
| 4.13 | Barabási e Albert e as Redes Livres de Escala | 87 |
| Capítulo 5. Rede Barabási-Albert Não-Linear | | 91 |
| 5.1 | A Distribuição de Conectividades Através da Equação Mestra | 94 |

| | | |
|--|--|-----|
| 5.2 | Distribuição de Conectividades na Aproximação de Campo Médio | 106 |
| 5.3 | Coeficiente de Agrupamento do Modelo BA não-linear | 108 |
| 5.4 | Coeficiente de Assortatividade | 112 |
| 5.5 | Comprimento Médio do Menor Caminho | 115 |
| 5.6 | Conclusões | 119 |
| Capítulo 6. A Rede Complexa do Futebol Brasileiro | | 121 |
| 6.1 | Física no Futebol | 121 |
| 6.2 | Rede Bipartida do Futebol Brasileiro | 122 |
| 6.3 | Rede Unipartida de Jogadores do Futebol Brasileiro | 127 |
| 6.4 | Comparação entre as Redes JFB, ER e de Configuração | 130 |
| 6.5 | Análise da Evolução Temporal da Rede JFB | 132 |
| Capítulo 7. Conclusões e Perspectivas | | 136 |
| Apêndices | | 145 |
| Apêndice A. Distribuições Lei de Potência, Zipf e Pareto | | 146 |
| Bibliografia | | 149 |

Lista de Figuras

| | | |
|-----|--|----|
| 2.1 | Pilha de areia. A inspiração para o conceito da criticalidade auto-organizada. | 7 |
| 2.2 | Avalanche de tamanho 6 numa pilha de areia. | 8 |
| 2.3 | Lei de Gutenberg-Richter para terremotos obtida no sistema estudado por Olami <i>et al.</i> [38]. | 11 |
| 3.1 | Informações coletadas pelo paleontólogo John Sepkoski Jr, ao longo de 10 anos [1]. | 14 |
| 3.2 | A distribuição de conectividade de roteadores. Rede de roteadores contendo 162 386 vértices, que possui uma conectividade média de 3,38. | 14 |
| 3.3 | Figura mostrando os passos da dinâmica co-evolutiva de Bak-Sneppen. . . . | 16 |
| 3.4 | As Figuras 3.4(a) e 3.4(b), retiradas do artigo [6], mostram as correlações espacial e temporal, respectivamente. | 17 |
| 3.5 | Exemplo de uma configuração do sistema vidro de <i>spin</i> que está armadilhada. Segundo o algoritmo EO o terceiro <i>spin</i> da rede sempre vai ser escolhido para alterar seu estado, fazendo com que o sistema fique preso uma configuração de mínimo local. | 20 |

| | | |
|------|---|----|
| 3.6 | Freqüência de adaptabilidade λ medida depois de alcançado o regime estacionário. Os círculos azuis correspondem ao algoritmo BS e exibe uma descontinuidade no ponto crítico $\lambda_c \sim 0,67$. As cruces verdes representam o algoritmo BSD com apenas 10 valores de adaptabilidade possíveis. As linhas tracejada, continua e tracejada com pontos correspondem ao algoritmo GEO-b para os valores de $\tau = 0,05, 0,50$ e $1,00$, respectivamente. . . | 27 |
| 3.7 | Adaptabilidade média versus τ . O algoritmo BS tem um valor de adaptabilidade média de $0,83$ que corresponde ao valor de $\tau = 0,86$ da dinâmica GEO-b , enquanto o valor de adaptabilidade média para o algoritmo Bak-Sneppen Discreto BSD é $0,86$ e corresponde a $\tau = 0,95$ da curva da dinâmica GEO-b . | 28 |
| 3.8 | A dinâmica GEO-b tem uma distribuição de probabilidade $C(x)$ constante que independe do valor de τ . O algoritmo BSD tem um comportamento misto enquanto o algoritmo BS segue uma lei de potência. | 29 |
| 3.9 | Distribuição de probabilidade de avalanches das dinâmicas BS e GEO-b . . . | 30 |
| 3.10 | Distribuições de probabilidades $C(x)$ para a dinâmica BSD com $Q = 10$, para valores de $\eta = 10^{-3}, 10^{-7}$ e 10^{-8} | 31 |
| 3.11 | Como já vimos na Figura 3.9 o algoritmo GEO-b tem uma distribuição de probabilidade $C(x)$ constante que independe do valor de τ . O algoritmo BSD tem um comportamento misto enquanto o algoritmo BSR , com um ruído (η) de apenas 10^{-3} , recupera o decaimento tipo lei de potência. | 32 |
| 4.1 | Pequena rede com oito vértices e dez conexões. | 35 |
| 4.2 | Figura 4.2(a) mostra um esboço das pontes de <i>Konigsberg</i> e o grafo correspondente é mostrado na Figura 4.2(b). | 37 |
| 4.3 | Histograma, mostrando que no final do experimento Milgram, as cartas que chegaram ao destino passaram por seis pessoas (em média). | 43 |

| | | |
|------|--|----|
| 4.4 | Rede bipartida, com 4 vértices do tipo filme e 11 vértices do tipo ator. Vértices de a a d representam filmes e os vértices de A até K representam atores, e as linhas ligam os atores aos filmes nos quais eles atuaram. | 46 |
| 4.5 | Rede unipartida, é a projeção da rede bipartida na qual atores se ligam uns aos outros se eles atuaram num mesmo filme. | 46 |
| 4.6 | Exemplo de dois tipos de redes: rede regular bidimensional 4.6(a) e rede aleatória 4.6(b). | 52 |
| 4.7 | Matriz adjacência com sua respectiva rede. | 54 |
| 4.8 | Densidade espectral de três ER, com $p = 0,05$ e tamanhos $N = 100$ (linha contínua), $N = 300$ (linha tracejada) e $N = 1000$ (linha pontilhada). [85] . . . | 55 |
| 4.9 | Rede social composta de três pessoas, onde A é amigo de B e de C . Nessas redes, existe uma grande chance de B ser amigo de C | 56 |
| 4.10 | Ilustração da definição do coeficiente de agrupamento (veja Equação 4.6). O coeficiente de agrupamento do vértice azul, localizado no centro da rede, para as três redes. Na primeira, o valor do coeficiente de agrupamento é zero , pois nenhum vizinho do vértice azul tem ligação entre si. Na segunda, quando existem duas ligações entre vizinhos do vértice azul, o coeficiente de agrupamento vale 1/3 e na terceira quando todos os vizinhos do vértice azul se conectam, o coeficiente de agrupamento é 1 , o valor máximo. | 57 |
| 4.11 | Rede com 19 vértices, utilizada para exemplificar o cálculo da correlação de conectividade $k_{nn,i}$ | 60 |
| 4.12 | Conectividade média dos primeiros vizinhos, $\bar{k}_{nn}(k)$, das redes de atores, de citações das áreas de astrofísica, física da matéria condensada e física de altas energias. Todas as quatro redes, seguem uma lei de potência com expoente $\nu \approx 0,3$ | 61 |

| | | |
|------|---|----|
| 4.13 | Conectividade média dos primeiros vizinhos das redes IR, AS e AS+. A linha sólida é dada pelo decaimento de lei de potência $\bar{k}_{nn}(k) \sim k^{-0,55}$, indicando que as redes AS e AS+ são correlacionadas e desassortativas. A linha horizontal pontilhada marca o valor da ausência de correlação $\bar{k}_{nn}(k) = \langle k^2 \rangle / \langle k \rangle = 26,9$, ou seja a rede IR é não assortativa. | 62 |
| 4.14 | Grafo com $N = 5$ vértices e 4 ligações. O conjunto de vértices é $P = \{1, 2, 3, 4, 5\}$, e o conjunto de ligações é $E = \{\{1, 2\}, \{1, 5\}, \{2, 3\}, \{2, 5\}\}$ | 63 |
| 4.15 | Representação da evolução do tamanho médio das componentes (linha contínua) e do tamanho da componente gigante (linha pontilhada) <i>vs.</i> a conectividade média, no caso de um grafo aleatório de Erdős-Rényi. | 67 |
| 4.16 | Construção de uma rede aleatória com uma dada distribuição de conectividades $P(k)$. (a) Conjunto de $\{k_i\}$ ligações incompletas. (b) Rede formada conectando aleatoriamente pares de $\{k_i\}$ ligações incompletas. | 68 |
| 4.17 | Representação das possíveis formas que os componentes de um grafo podem assumir. | 77 |

| | | |
|------|--|----|
| 4.18 | <p>Procedimento de reconectar os vértices do modelo WS que transforma uma rede (circular) regular em uma rede aleatória sem fazer mudança no número de vértice ou de ligação. Começamos com uma rede de $L = 16$ vértices, cada um conectado aos seus 4 vizinhos mais próximos, em seguida escolhemos um vértice e uma ligação que conecta um de seus vizinhos. Com uma probabilidade p reconectamos essa ligação a um outro vértice escolhido aleatoriamente, fazemos isso até que todas as ligações, da rede original, tenham sido considerados. Realizações deste procedimento é mostrado, aqui nesta figura, para três valores diferentes de p. Na primeira, foi usado $p = 0$, com isso a rede original permanece inalterada. Na segunda, com um valor de p no intervalo $0 \leq p \leq 1$. Na terceira, é obtida uma rede totalmente aleatória com $p = 1$, onde todas as ligações foram aleatoriamente reconectadas.</p> | 85 |
| 4.19 | <p>Coeficiente de agrupamento $C(p)$ e o comprimento médio dos menores caminhos $\ell(p)$ do modelo Watts-Strogatz. Os dados foram normalizados pelos valores de $C(0)$ e $\ell(0)$, respectivamente. Pode ser visto a rápida queda de $\ell(p)$, demonstrando que a rede se comporta como uma rede mundo pequeno. Durante essa queda de $\ell(p)$, $C(p)$ permanece praticamente constante, o que indica que a transição para uma rede mundo pequeno é quase imperceptível localmente.</p> | 86 |
| 4.20 | <p>Distribuição de conectividades de redes World-Wide Web de diferentes tamanhos. Os quadrados correspondem a uma amostra de 325.729 de páginas (obtidos por Réka Albert) e os círculos correspondem a uma medida de 200 milhões de páginas (obtidos por Broder et al.). (a) Distribuição de conectividades de ligações que recebem ligações de outras páginas. (b) Distribuição de conectividades de ligações que enviam ligações para outras páginas. . . .</p> | 88 |

| | | |
|------|--|-----|
| 4.21 | A distribuição de conectividades de uma rede aleatória, com $N = 10.000$ vértices e probabilidade de conexão $p = 0,0015$. A linha contínua mostra o ajuste da distribuição de conectividades segundo uma distribuição de Poisson. | 89 |
| 5.1 | Distribuição de conectividades do modelo BA , com $N = m + t = 300\ 000$ e $m = 2$). | 93 |
| 5.2 | Distribuição de conectividades de uma rede BA-não linear para o caso de $\alpha = 1$ e $m = 4$. Os quadrados vermelhos são pontos que foram obtidos via simulação numérica, enquanto os círculos azuis foram obtidos via cálculo analítico Eq. (5.27) | 100 |
| 5.3 | Comportamento de μ em função de α | 102 |
| 5.4 | Evolução temporal de $\theta(s)$ para os instantes de tempo $s = 1, 2, 3$, respectivamente. | 105 |
| 5.5 | Rede na vizinhança do vértice i (rede à esquerda). Rede depois de removido o vértice i e suas ligações (rede à direita). | 109 |
| 5.6 | O coeficiente de agrupamento médio C vs o expoente α para dois diferentes tamanhos de rede. | 112 |
| 5.7 | Comparação entre simulação numérica e a teoria. Quadrados pretos são os pontos de simulação e os círculos abertos correspondem a formula analítica proposta. | 113 |
| 5.8 | A dependência do coeficiente de assortatividade r com o expoente α | 114 |
| 5.9 | O mínimo comprimento médio do menor caminho ℓ_{min} vs o expoente α para dois diferentes tamanhos de rede. | 117 |
| 5.10 | O modelo BA linear ($\alpha = 1$). Gráfico mostrando a dependência do mínimo comprimento médio do menor caminho ℓ_{min} vs $\ln(N)$ e $\ln(\ln(N))$ para $m = 1$ e $m = 5$, respectivamente. | 118 |

| | | |
|-----|---|-----|
| 6.1 | Rede bipartida de futebol, com 4 vértices do tipo clube e 11 vértices do tipo jogador de futebol. Vértices de a a d representam clubes e os vértices de A até K representam jogadores, e as linhas ligam os jogadores aos clubes que defenderam. | 122 |
| 6.2 | Histograma do número de clubes contra o número de gols marcados por partida. A linha cheia corresponde ao ajuste gaussiano. O número médio de gols por partida é aproximadamente 1,0. | 123 |
| 6.3 | Probabilidade $P(N)$ de que um jogador tenha trabalho em N clubes. A linha cheia corresponde ao ajuste da curva $P(N) \sim 10^{-0,38N}$. Assim sendo, vemos que é 190 vezes mais provável de encontrar algum jogador que tenha jogado somente em dois clubes do que encontrar um jogador que tenha jogado em oito clubes. No gráfico menor vemos a distribuição de probabilidade de conectividade $P(S)$ dos clubes. | 125 |
| 6.4 | (a) Probabilidade de jogos $P(M)$ vs o número M de partidas disputadas. (b) Distribuição de probabilidade cumulativa $P_c(M)$ construída a partir de $P(M)$. As linhas cheias são os ajustes dos conjuntos de pontos. | 126 |
| 6.5 | (a) Probabilidade $P(G)$ que um jogador tenha marcado G gols. É dez vezes mais provável que um jogador escolhido aleatoriamente tenha marcado 13 gols do que 36 gols. O jogador com o maior número de gols marcados no campeonato brasileiro é o Roberto Dinamite com $G = 186$. (b) Distribuição de probabilidade cumulativa $P_c(G)$ | 127 |
| 6.6 | Rede unipartida, é a projeção da rede bipartida na qual os jogadores se ligam uns aos outros se eles defenderam um mesmo clube. | 128 |
| 6.7 | Distribuição de conectividades da rede de jogadores do futebol brasileiro, onde um jogador se conecta a outro se eles defenderam um <i>mesmo clube</i> num <i>mesmo ano</i> . A curva de ajuste (linha cheia) tem a forma exponencial $P(k) \sim e^{-0,025k}$ | 129 |

| | | |
|------|---|-----|
| 6.8 | Distribuição de conectividades de atores de filmes de Hollywood. Essa figura foi retirado do artigo pela Réka Albert e Albert-László Barabási intitulado <i>Topology of Evolving Networks: Local Events and Universality</i> (publicado na revista <i>Physical Review Letters</i> 85 , 5234). | 130 |
| 6.9 | Conectividade Média dos Primeiros Vizinhos das redes de Configuração, JFB e ER | 133 |
| 6.10 | Diagrama que mostra a rede de transferências entre os clubes do futebol brasileiro. O número dentro do círculo representa o número de clubes com o qual cada clube já negociou (comprou/vendeu/emprestou) jogadores. O Bahia (circulo em vermelho, no centro da figura) é o clube com maior número de transferências, 88. | 135 |
| A.1 | Ranking de palavras da língua inglesa. A curva mostra que algumas palavras são muito mais freqüentes do que outras. Um exemplo é a palavra “the” que é quase 1000 vezes mais freqüente do que a palavra “quality”. | 147 |

Lista de Tabelas

| | | |
|-----|---|-----|
| 3.1 | Lista dos valores de adaptabilidade crítica, λ_c , para redes regulares de dimensões $\mathbf{d} = 1, 2, 3, 4, 5$ e 6 | 18 |
| 4.1 | Definição de alguns dos termos mais utilizados no estudo das redes. | 48 |
| 4.2 | As propriedade medidas são: número total de vértices N ; número total de ligações m ; conectividade média z ; distância média entre vértices ℓ ; coeficiente de agrupamento C . Na última coluna estão algumas referências de onde os dados foram obtidos. | 83 |
| 6.1 | Na primeira coluna, \mathbf{v} é o número de vértices, \mathbf{e} é o número de ligações, $\bar{\mathbf{k}}$ é a conectividade média, \mathbf{C} é o coeficiente de agrupamento, \mathbf{A} é o coeficiente de assortatividade e $\bar{\mathbf{D}}$ é o comprimento médio do menor caminho. | 132 |
| 6.2 | Evolução temporal de algumas quantidade da rede JFB . O significado das grandezas na primeira coluna são os mesmos da Tabela 6.1. | 134 |

CAPÍTULO 1

Introdução

Informatio simplex procul scisco res rei difficilis

Idéias simples para resolver coisas complicadas

O homem sempre desejou entender o mundo à sua volta, primeiramente em razão da sua própria sobrevivência. Rodeado por forças materiais, o primeiro objetivo do homem foi dominar o seu meio ambiente. A segurança contra predadores e fenômenos naturais, a busca da alimentação, a organização social dos núcleos humanos etc., despertaram os primeiros questionamentos do homem sobre as leis que governam a natureza. Na medida em que as necessidades do espírito humano foram se tornando mais complexas, cresceu também a necessidade de uma melhor compreensão do mundo. Devido a enorme complexidade do mundo o homem passou a formular modelos simples que representassem, da melhor maneira possível, as forças interativas que caracterizam um fenômeno.

Um modelo é um veículo para *uma visão bem estruturada da realidade*. Um modelo também pode ser visto, com os devidos cuidados, *como uma representação substitutiva da realidade*. Lidando com o concreto ou abordando o imaginário, a mente do

homem trabalha com estruturas de “substituição” que visam facilitar o raciocínio. A essas estruturas denominamos genericamente de modelos.

É importante que seja dito que, para que sejam *implementáveis*, os modelos devem ser, preferencialmente, livres de pequenos detalhes onerosos. Sendo assim, o equilíbrio entre a simplicidade e a validade, é fundamental.

Seguindo essa linha de raciocínio em 1987, Per Bak, Chao Tang e Kurt Wiesenfeld [1] propuseram o modelo da pilha de areia, que é uma simplificação do fenômeno de avalanches. A dinâmica deste modelo consiste em adicionar grãos de areia numa mesa até a formação de uma pilha que, com o passar do tempo passa a ter avalanches. Com o auxílio desse experimento, foi possível constatar que a pilha de areia exhibe o comportamento de equilíbrio entrecortado (*punctuated equilibrium*), onde períodos de *quiescência* são interrompidos por intermitentes deslizamentos dos grão de areia. Os deslizamentos dos grãos, ou avalanches, são causados por reações em cadeia, nas quais um simples grão de areia pode levar consigo um ou mais grãos.

Segundo Per Bak, os sistemas críticos auto-organizados evoluem para um estado complexo crítico sem a interferência de qualquer agente externo. O processo de auto-organização acontece depois de um período muito longo de transiente. Comportamentos complexos, seja na geofísica ou na biologia, emergem sempre depois de um longo processo de evolução. Em outras palavras, não podemos querer entender um determinado sistema, olhando apenas para um pequeno intervalo temporal. Dessa forma, a frase “não podemos entender o presente sem antes entender o passado” ganha um significado mais preciso e profundo.

A natureza é composta de seres vivos e matéria inerte que evoluem *fora do equilíbrio*. Isto pode ser evidenciado medindo-se escalas de tempo apropriadas. Em particular, a dinâmica dos seres vivos é basicamente governada pela teoria da seleção natural, que foi proposta por Charles Darwin [2, 3]. Com certeza, esta teoria foi uma das grandes

realizações do século XIX. A teoria da seleção natural diz que a evolução não se dá expressamente pela procriação das espécies melhores adaptadas ao seu meio, mas sim pela extinção das espécies menos adaptadas. Este mecanismo leva à emergência de estruturas altamente especializadas.

Baseado nas idéias Darwinianas, Per Bak e Kim Sneppen propuseram um modelo de evolução biológica, simples e robusto, para simular a dinâmica da co-evolução das espécies.

No Capítulo 3, apresentamos um estudo comparativo entre o modelo Bak-Sneppen (**BS**) [4–9] e algoritmos de otimização [10–13], com vários modelos que propusemos e que modificam a dinâmica **BS** padrão.

Depois de termos estudado algoritmos de otimização, durante todo o mestrado e início do doutorado, nos voltamos para o estudo das redes complexas. Assim como a criticalidade auto-organizada, o conceito de rede complexa tem-se mostrado quase que onipresente na natureza.

Nosso interesse em redes complexas começou quando buscávamos um algoritmo de construção de uma rede que pudesse gerar redes de diferentes topologias. Foi então que nos deparamos com o artigo intitulado “The Bak-Sneppen Model on Scale-Free Networks”, dos autores Yamir Moreno e Alexei Vazquez [14], onde estudam a criticalidade auto-organizada numa rede complexa [15–18].

O estudo das redes complexas, que começou com o nascimento da teoria de grafos, hoje permeia e até une várias áreas, tornando a interdisciplinaridade uma realidade na pesquisa científica de vanguarda.

Nos últimos anos, várias redes reais têm sido estudadas pela comunidade científica, sendo que algumas delas se tornaram famosas. Em seu artigo de revisão das redes complexas, Mark Newman [19], propõe a classificação das redes complexas da seguinte forma: tecnológicas, biológicas, sociais e de informação. O exemplo mais conhecido das redes tecnológicas, sem dúvida alguma, é a rede da Internet [20–22]. No grupo das

redes biológicas, temos o exemplo clássico da rede de interação proteína-proteína [23]. Dentre as redes de informação, também chamadas de redes de conhecimento, um dos exemplos mais conhecidos é o da rede de citações de artigos científicos [24]. Bons exemplos de redes sociais são: os de co-autoria em artigos científicos [25–29], de contatos sexuais [30], a rede bipartida atores-filmes de Hollywood [31, 32] e, finalmente, o da rede de jogadores do futebol brasileiro [33], que foi um dos objetos de estudo desta tese.

Diferentes grandezas ou parâmetros são utilizados na descrição e/ou caracterização das redes complexas. Está disponível na rede Internet uma lista bastante completa dessas grandezas [34]. Esta lista vêm sendo sistemática e periodicamente atualizada. Talvez a grandeza mais importante seja a distribuição de conectividade, que é definida como sendo a fração de vértices da rede que tem conectividade k , ou ainda, em outras palavras como a probabilidade de que um vértice escolhido ao acaso tenha conectividade k . Temos também a grandeza conhecida como comprimento médio do menor caminho, que é obtido calculando-se o valor médio dos caminhos mais curtos (geodésicas) entre todos pares de vértices da rede. Outra grandeza muito utilizada na caracterização das redes é o coeficiente de agrupamento, que é obtido calculando-se o valor médio das probabilidades de que os primeiros vizinhos de um determinado vértice também estejam conectados entre si. Uma outra grandeza, que também recebeu nossa atenção nessa tese, foi o coeficiente de assortatividade. Essa grandeza mede a tendência dos vértices de uma rede de se conectarem, preferencialmente, com outros vértices de mesma conectividade ou de conectividade distinta.

Redes complexas — o que são e para que elas servem — é o assunto principal do Capítulo 4, onde apresentamos uma breve revisão histórica e conceitual. No Capítulo 5 estudamos detalhadamente a rede Barabási-Albert não-linear. No Capítulo 6 analisamos a rede complexa do futebol brasileiro. E, por fim no Capítulo 7 apresentamos nossas conclusões e perspectivas.

CAPÍTULO 2

Breve Introdução a Criticalidade Auto-Organizada

2.1 O Todo é Maior do que a Soma das Partes

*D*urante as últimas duas décadas os sistemas complexos têm despontado como um tema interdisciplinar que conecta áreas tradicionais como física, química, biologia, ecologia e até mesmo ciências sociais tais como economia e sociologia. Sistemas podem ser considerados complexos quando mostram emergência de formas coerentes longe da aleatoriedade. Uma definição mais compreensível é a de que sistemas complexos são compostos de um grande número de componentes ou “agentes”, interagindo de tal modo que o comportamento coletivo não é uma simples combinação dos comportamentos individuais.

A principal propriedade dos sistemas complexos reside na possibilidade de ocorrer uma ação coletiva e coerente em grande escala, produzindo estruturas (padrões) interessantes. Um exemplo muito curioso são as manchas do leopardo, resultado de repetidas interações entre componentes que constituem seu pêlo, levando o estudo do sistema como um todo a ser mais interessante do que o estudo de uma ou outra parte em separado.

Eventos como furacões, incêndios florestais, terremotos e avalanches são exemplos de sistemas complexos que se organizam por meio de uma dinâmica de interação não-linear.

2.2 Criticalidade Auto-Organizada

O termo “*self-organized criticality*” (SOC) ou criticalidade auto-organizada, como foi traduzido para o português, refere-se à tendência de um sistema dinâmico a organizar-se num estado estacionário, sem qualquer parâmetro externo de ajuste [9]. Outras propriedades importantes do estado crítico auto-organizado são: a existência de um longo período transiente, as correlações espacial e temporal exibem escala tipo lei de potência e há grandes instabilidades denominadas avalanches. Neste caso, longos períodos de quiescência são seguidos de frenética atividade — a isso chamamos de equilíbrio entrecortado ou pontuado (do inglês, *punctuated equilibrium*).

O estudo de dinâmicas fora do equilíbrio em sistemas heterogêneos, como o modelo de pilha de areia (que veremos com mais detalhes na seção 2.3) e o modelo co-evolutivo de Bak-Sneppen, demonstram a importância de uma abordagem metódica, onde sistemas com um grande número de partes interagindo mutuamente, trocando energia, matéria ou informação com o meio, auto-organizam suas estruturas internas e dinâmicas com novas e, algumas vezes, surpreendentes propriedades macroscópicas.

2.3 Modelo da Pilha de Areia

O conceito de criticalidade auto-organizada foi introduzido por Per Bak, Chao Tang e Kurt Wiesenfeld [4] em 1987 tendo como inspiração uma pilha de areia (veja a Figura 2.1).

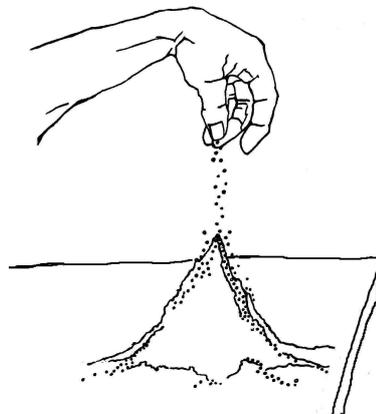


Figura 2.1: Pilha de areia. A inspiração para o conceito da criticalidade auto-organizada.

A dinâmica Bak-Tang-Wiesenfeld (**BTW**) [7], reproduz o crescimento de uma pilha de areia. No caso de uma rede quadrada consiste dos seguintes passos: Cada sítio receberá uma quantidade finita de grãos, $z(x, y)$ (altura da coluna de areia na posição (x, y)), variando de 0 (zero) até $z_c(x, y)$, onde $z_c(x, y)$ é um valor crítico a partir do qual ocorre um deslizamento. Para o exemplo que apresentaremos a seguir adotamos $z_c(x, y) = 4$. Logo, cada sítio poderá conter $0 \leq z(x, y) \leq 3$ grãos de areia sem que ocorra um deslizamento. A cada passo temporal adicionamos um novo grão de areia num sítio, que é escolhido aleatoriamente. O valor de $z(x, y)$ desse sítio é acrescido de 1. Esse procedimento se repete várias vezes até que algum $z(x, y)$ atinja o valor 4. Haverá, então, um deslizamento. Quando este deslizamento ocorre, o seguinte procedimento é implementado:

$$\begin{aligned}
 z(x, y) &\rightarrow 0 \\
 z(x \pm 1, y) &\rightarrow z(x \pm 1, y) + 1 \\
 z(x, y \pm 1) &\rightarrow z(x, y \pm 1) + 1
 \end{aligned}
 \tag{2.1}$$

Isto significa que os quatro grãos que estavam no sítio (x, y) são distribuídos entre os quatro primeiros vizinhos. A distribuição é feita em partes iguais, ou seja, cada vizinho recebe um grão de areia do sítio central. Caso algum sítio receba um grão de um vizinho e complete quatro grãos, o processo se repetirá. Conseqüentemente, haverá um novo deslizamento e uma redistribuição de grãos, podendo resultar em uma reação em cadeia de longo alcance, isto é, uma avalanche. Uma avalanche será caracterizada pelo seu tamanho, que é contabilizado pelo número de sucessivos deslizamentos que acontecem desde o momento em que foi acrescentado o último grão. Por exemplo, se ao adicionarmos um grão, na rede, este gerar um único deslizamento, dizemos que ocorreu uma

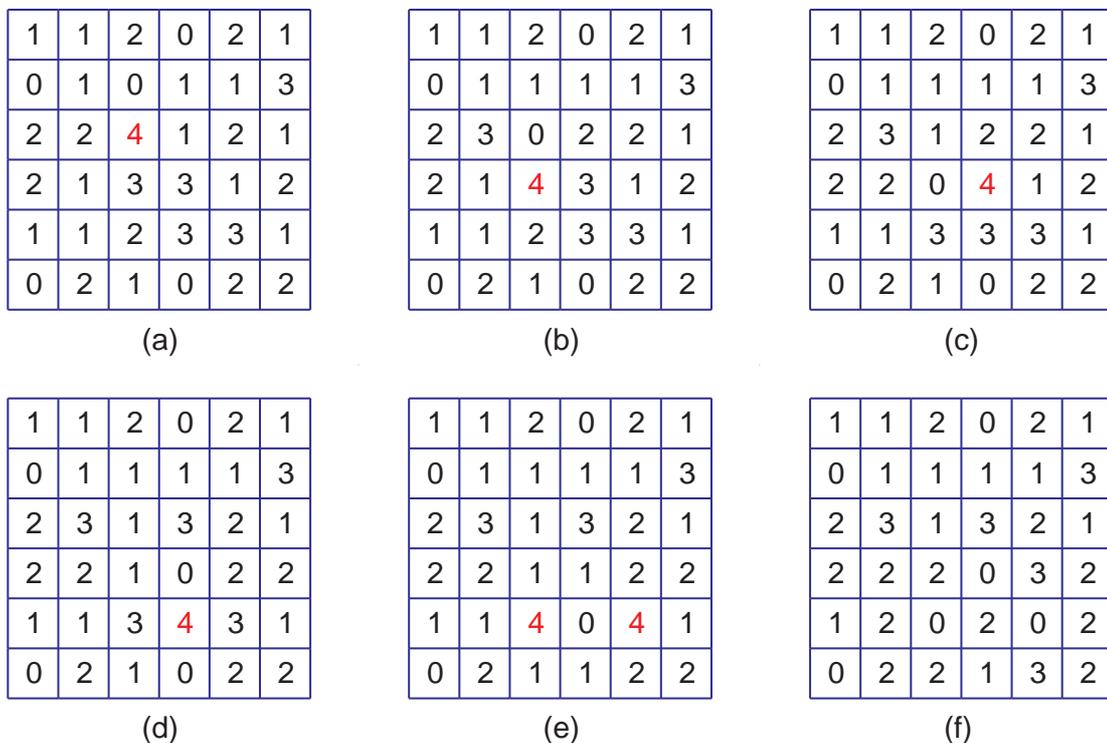


Figura 2.2: Avalanche de tamanho 6 numa pilha de areia.

avalanche de tamanho um, se o número de deslizamentos gerados for dois, então concluímos que ocorreu uma avalanche de tamanho dois, e assim por diante.

Na Figura 2.3 [35]¹, temos uma certa distribuição inicial de grãos de areia. Deixamos então cair um grão de areia na célula $z(3, 3)$ que completa a condição estabelecida para a ocorrência de um deslizamento. Haverá então uma redistribuição dos grãos deste sítio para os seus primeiros vizinhos. Estes podem eventualmente também atingir o valor crítico e realizar o mesmo processo anterior. Na Figura 2.2 é mostrada uma avalanche de tamanho 6. Conforme a dinâmica evolui, avalanches de tamanhos do mais variados tamanhos podem ocorrer.

Vale lembrar que o modelo **BTW** é uma simplificação da pilha de areia real, visto que os grãos apresentam formas e propriedades diversas entre eles, e o ambiente também pode contribuir com a gravidade, resistência do ar, etc.

Experimentos reais com pilha de areia feitos em laboratório [36, 37] não apresentaram evidências de criticalidade devido ao efeito de inércia, uma vez que o grão da areia encontra dificuldade para superar o ângulo de repouso. Contudo, isso foi facilmente contornado, utilizando-se grãos de arroz, ao invés de areia, uma vez que grãos de arroz têm uma forma alongada e com isso o efeito de inércia é minimizado.

Outro modelo muito estudado, que também apresenta criticalidade auto-organizada, é o modelo Olami-Feder-Christensen (**OFC**) [38–40], que pode ou não ser conservativo. O modelo **OFC** é uma versão simplificada do modelo de terremotos de Burridge-Knopoff [41]. Numa rede unidimensional com L sítios, o modelo **OFC** pode ser definido da seguinte forma:

- i) para cada sítio i , é atribuída uma variável z_i que pode assumir valores entre $0 \leq z_i \leq z_c$ representando uma energia (ou tensão) local. z_c é o valor crítico dessa energia;
- ii) se em algum sítio $z_i \geq z_c$ então a energia z_i é distribuída entre seus vizinhos, de acordo com a regra

¹software livre de processamento de redes Pajek que pode ser baixado do site <http://vlado.fmf.uni-lj.si/pub/networks/pajek/>

$$z_i \rightarrow 0$$

e

$$z_{nn} \rightarrow z_{nn} + \alpha z_i$$

onde z_{nn} são todos os primeiros vizinhos do vértice i e o parâmetro α , que tem seu valor entre 0 e 0,5, controla o nível de conservação do modelo. No caso da rede unidimensional, quando α é igual a 0,5 o sistema é dito conservativo, uma vez que a energia perdida pelo vértice i é redistribuída de forma conservativa entre seus primeiros vizinhos;

Nesse instante um terremoto está acontecendo.

- iii) repita o passo ii) até o terremoto terminar;
- iv) faça $z_i \rightarrow z_i + \Delta z_i$ para cada vértice da rede (perturbação global), onde Δz_i é um valor de incremento, e retorne ao passo ii).

Concluída a dinâmica **OFC**, podemos medir a distribuição de probabilidade dos tamanhos de terremotos, que é proporcional a energia liberada durante um terremoto.

Vale lembrar que os modelo **BTW** e **OFC** representam uma classe de sistemas que são conduzidos para um estado estacionário caracterizado por funções de correlação espacial e temporal decaindo como uma lei de potência. Ou seja, eventos de todos os tamanhos (ou intensidade) têm probabilidade de ocorrência, não havendo um tamanho característico no qual o sistema estaria mais propenso a operar.

Em 1956, Beno Gutenberg e Charles Francis Richter [42] verificaram que o número de terremotos de magnitude M se relaciona com a energia E liberada através da lei de potência

$$P(E_0 - E) \sim E^{-B}, \quad (2.2)$$

essa é a notória lei de Gutenberg-Richter (veja Figura 2.3), onde E_0 é um valor de energia de referência e o expoente B assume valores positivos.

No Capítulo 3, apresentaremos o modelo co-evolutivo de Bak-Sneppen **BS** e os modelos evolutivos Otimização Extrema **EO** e Otimização Extrema Generalizada **GEO**, que foram propostas com a intenção de capturar a dinâmica co-evolutiva e evolutiva de espécies, respectivamente. Apresentaremos ainda dois outros modelos, que foram propostos por nós [43], o Bak-Sneppen Discreto (**BSD**) e o Bak-Sneppen com Ruído (**BSR**), para estudar a influência e importância do papel da biodiversidade na co-evolução de espécies. Ainda no capítulo 3, faremos um estudo comparativo entre as dinâmicas **BS**, **BSD**, **BSR** e **GEO**.

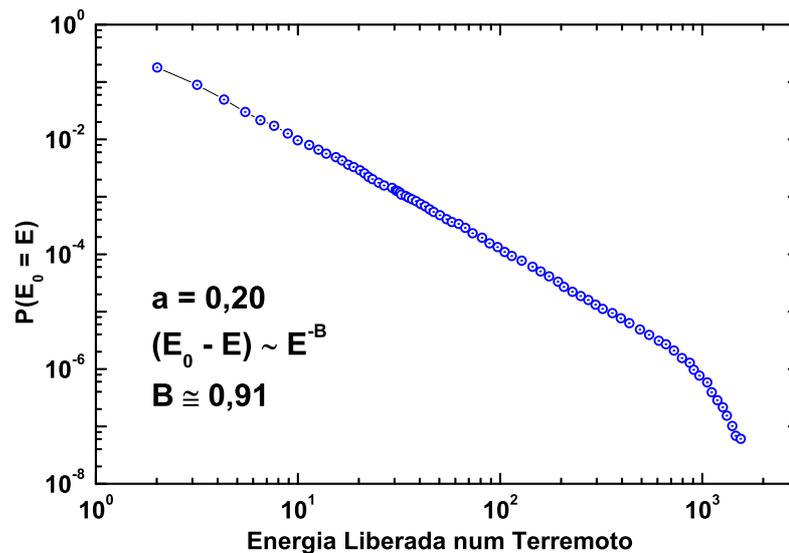


Figura 2.3: Lei de Gutenberg-Richter para terremotos obtida no sistema estudado por Olami *et al.* [38].

CAPÍTULO 3

Criticalidade Auto-Organizada, Otimização e Biodiversidade

3.1 Sistemas Complexos

*U*ma das áreas da física que lida com comportamento coletivo de elementos é a mecânica estatística. Um dos objetivos dessa disciplina, dentre tantos outros, é estudar transições de fases — quando o comportamento coletivo é alterado devido a mudança do parâmetro de ordem, por exemplo mudança de condições do meio ambiente como temperatura, pressão, etc.. Efeitos similares em sistemas complexos na biologia [44] e ciências sociais [45] têm sido estudados, auxiliando na identificação das características mínimas para que ocorra a emergência de comportamento coletivo.

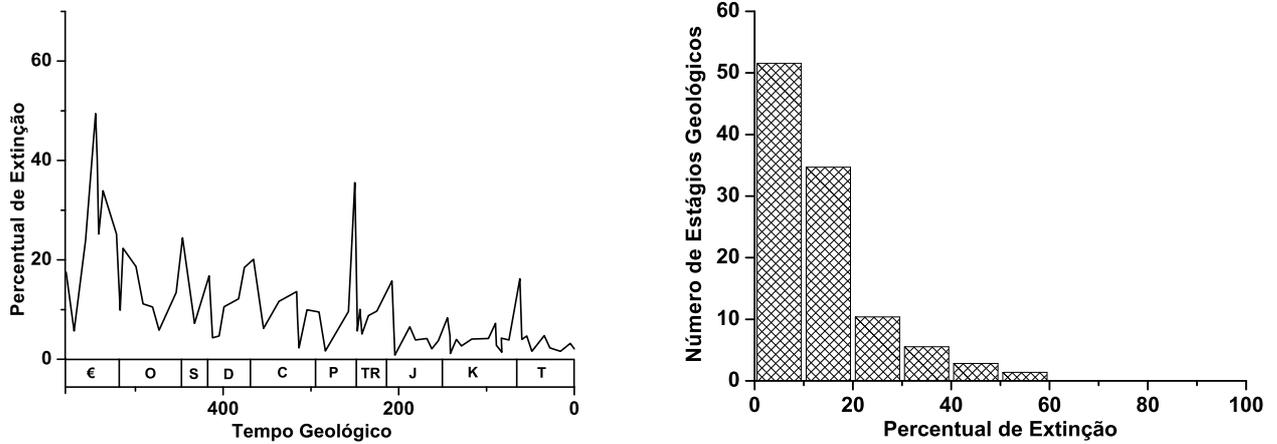
Um dos maiores entusiastas do estudo de sistemas complexos foi Per Bak. Além disso é creditado a ele a proposta de um processo, que ajuda a melhor compreender os sistemas complexos, chamado criticalidade auto-organizada — auto-organizada porque não há nenhum parâmetro externo operando sobre o sistema, e é crítico porque ele se equilibra no ponto crítico entre a ordem e a desordem. Isto é, onde as coisas interessantes acontecem. Como é o caso da vida.

Uma das características fundamentais de um sistema no estado crítico auto-organizado é a de que, no estado estacionário, as correlações espacial e temporal decaem seguindo leis de potência. Com grande frequência encontramos fenômenos e sistemas na natureza regidos por uma lei de potência. Um exemplo de sistema natural regido por lei de potência são os eventos de extinção. Na década de 1980, os paleobiólogos Jack John Sepkoski Jr. e David M. Raup [46], com objetivo de estudar os eventos de extinção, fizeram um rigoroso levantamento e análise estatística de dados de fósseis marinhos, levantamento esse que Sepkoski demorou 10 anos para concluir.

Sepkoski dividiu a história geológica em períodos consecutivos de alguns milhões de anos. Para cada período, ele estimou a fração de espécies que havia se extinguido (veja a Figura 3.1(a)). Aconteceram extinções pequenas, que exterminaram menos de 5 por cento das espécies, e outras maiores que envolveram até 50 por cento das espécies do planeta. O famoso período Cretáceo no qual foram extintos os dinossauros não está entre os mais proeminentes. Numa outra análise, Raup contou o número de períodos nos quais o número de extinções era menor do que 10 por cento, quantos períodos apresentavam um número de extinção entre 10 e 20 por cento e assim por diante (veja a Figura 3.1(b)).

Assim como os sistemas naturais, alguns sistemas construídos pelo homem também são regidos por lei de potência, um bom exemplo é a rede Internet. Faloutsos et al. [20] estudaram a rede Internet¹ e concluíram que ambas as distribuições de conectividade, de roteadores e de domínios, decaíam como uma lei de potência, como podemos ver na Figura 3.2.

¹no caso da rede de roteadores, os vértices são os roteadores e as ligações são as conexões entre os roteadores, e no caso da rede de domínios, os vértices são os domínios, e dois domínios estão ligados se existir pelo menos um caminho entre eles.



(a) Eventos de extinções biológicas ocorridas nos últimos 600 milhões de anos. (b) Histograma dos eventos de extinção da Figura 3.1(a).

Figura 3.1: Informações coletadas pelo paleontólogo John Sepkoski Jr, ao longo de 10 anos [1].

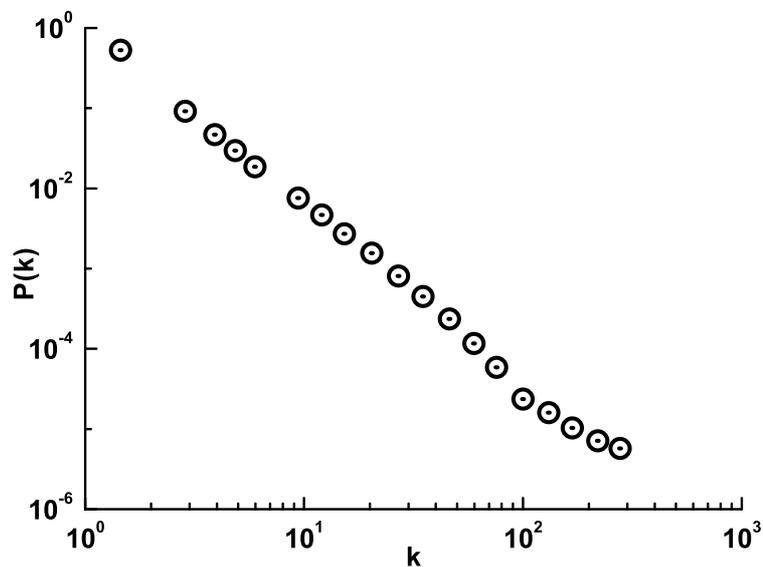


Figura 3.2: A distribuição de conectividade de roteadores. Rede de roteadores contendo 162 386 vértices, que possui uma conectividade média de 3,38.

3.2 Dinâmicas de Otimização

Per Bak e Kim Sneppen [6] propuseram um modelo de co-evolução biológica, simples e robusto, para simular a dinâmica de espécies, que ficou conhecido como

modelo de Bak-Sneppen (**BS**). A simplicidade da dinâmica começa pelo fato de toda uma espécie ser representada por um simples valor de adaptabilidade (*fitness*). A adaptabilidade de cada espécie é afetada pela presença de espécies vizinhas. Apesar da simplicidade, o modelo apresenta o fenômeno do equilíbrio entrecortado.

Inspirado na dinâmica **BS**, Stefan Boettcher e Allon G. Percus [10, 11], propuseram a dinâmica **EO** e sua forma modificada a dinâmica **GEO**.

Nas Seções 3.2.1 e 3.2.2 deste Capítulo apresentamos os algoritmos de otimização **BS**, **EO** e **GEO**, onde o primeiro algoritmo possui característica co-evolucionária enquanto os outros dois possuem característica evolucionária. A partir da Seção 3.3.1 estudamos a eficiência dos algoritmos, por meio de uma análise comparativa dos valores de adaptabilidade no regime estacionário, das dinâmicas **BS** (e algumas de suas variantes) e **GEO**. Apresentamos ainda, um estudo das correlações espacial e temporal de todas as dinâmicas envolvidas.

3.2.1 Modelo Co-Evolutivo de Bak-Sneppen

Um dos paradigmas da criticalidade auto-organizada é o modelo co-evolutivo de Bak-Sneppen (**BS**) [6]. Nesse modelo, diferentes espécies num mesmo ecossistema se relacionam, por exemplo, através da cadeia alimentar e da co-evolução guiada pelos princípios Darwinianos, onde a mutação e a extinção de uma espécie afeta a evolução de espécies vizinhas. O algoritmo proposto é o seguinte:

- i) Colocar N espécies, uma em cada sítio, de uma rede unidimensional com condições periódicas de contorno (veja a Figura 3.3(a));

- ii) Atribuir um valor de adaptabilidade, λ_i (sorteado de uma distribuição uniformemente distribuída entre 0 e 1), para cada uma das espécies i ;
- iii) Localizar a espécie com menor valor de λ_i ;
- iv) Atribuir a espécie com menor valor de adaptabilidade e a seus primeiros vizinhos novos valores de λ_i , também sorteados de uma distribuição uniformemente distribuída (veja a Figura 3.3(b));
- v) Repetir os passos (iii), (iv) e (v) o quanto se julgar necessário.

No exemplo da Figura 3.3(a), vemos os passos i) e ii) implementados, em que há a atribuição dos seguintes valores de adaptabilidade a cada uma das 6 espécies: $\lambda_1 = 0,67$, $\lambda_2 = 0,83$, $\lambda_3 = 0,58$, $\lambda_4 = 0,08$, $\lambda_5 = 0,32$ e $\lambda_6 = 0,21$. Na Figura 3.3(b) vemos o passo iii) da dinâmica de co-evolução **BS**, onde o vértice 4, por possuir o menor valor de adaptabilidade $\lambda_4 = 0,08$ é selecionado, para ter seu valor de adaptabilidade atualizado junto com os vértices 3 e 5.

Per Bak e Kim Sneppen, estudando a dinâmica co-evolutiva de espécies sob a luz do modelo **BS**, encontraram os seguintes resultados (veja Figura 3.4).

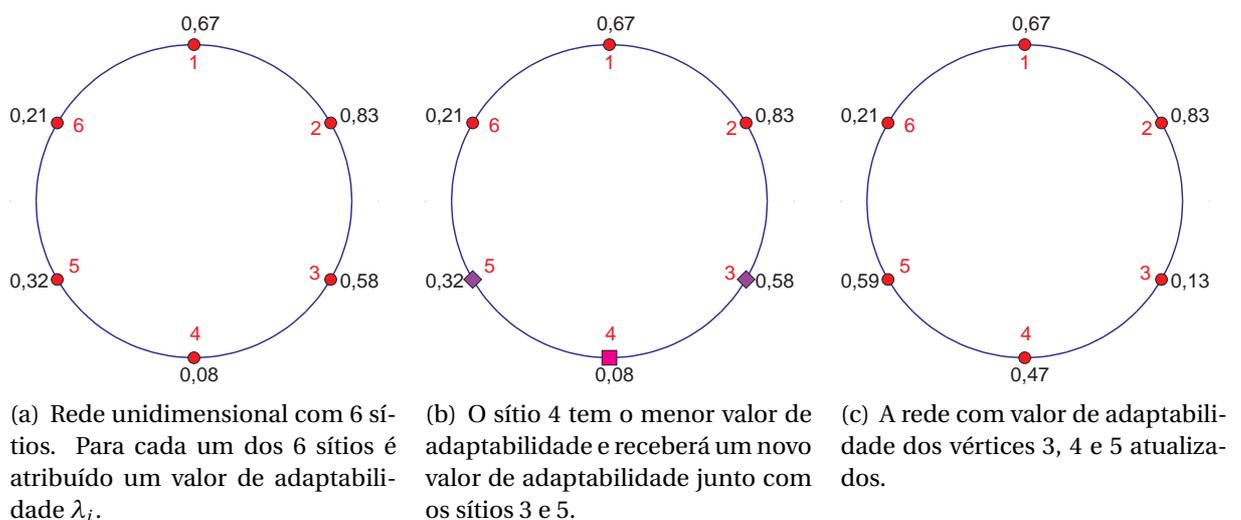
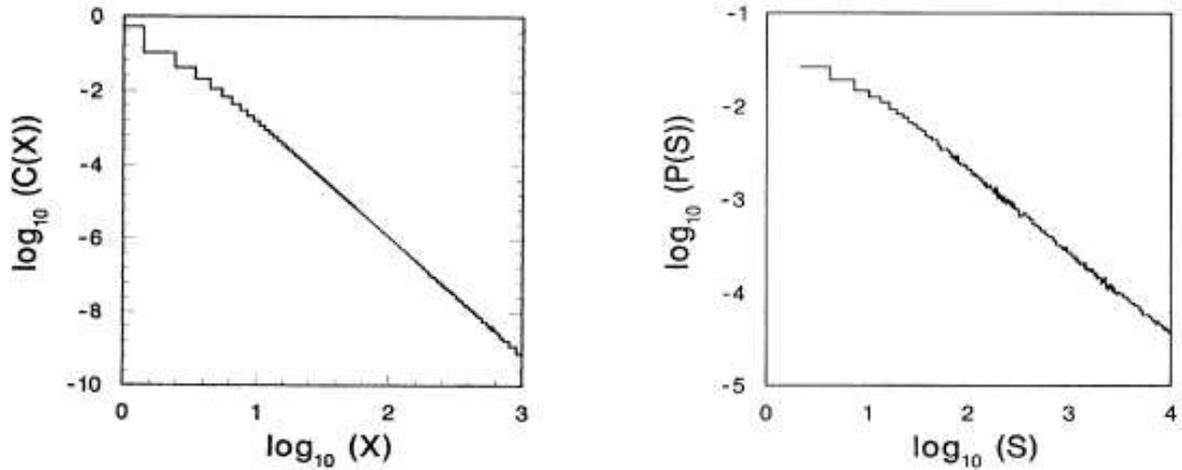


Figura 3.3: Figura mostrando os passos da dinâmica co-evolutiva de Bak-Sneppen.



(a) Probabilidade de distribuição de distâncias entre sucessivas mutações.

(b) Distribuição de avalanches no estado crítico.

Figura 3.4: As Figuras 3.4(a) e 3.4(b), retiradas do artigo [6], mostram as correlações espacial e temporal, respectivamente.

Chamamos de X a distância geométrica entre dois subseqüentes sítios ativos. São ativos, os sítios selecionados de menor valor de adaptabilidade. Na Figura 3.4(a) vemos a medida da probabilidade de distribuição de distâncias $C(X)$, onde $C(X)$ é o número de vezes (frequência) em que o algoritmo **BS** encontra a distância X como sendo a separação entre dois subseqüentes sítios ativos.

A distribuição de distâncias obedece a lei de potência, $C(X) = X^{-3,15 \pm 0,05}$ [6], mostrando que o sistema é crítico. Isso não depende de nenhuma condição inicial, então o estado crítico é um atrator global para a dinâmica, e conseqüentemente podemos dizer que o sistema é auto-organizado.

Denominamos S como sendo o tamanho de uma avalanche, que é definido pelo número de mutações (atividades) sucessivas abaixo de um valor de adaptabilidade previamente estabelecido $\bar{\lambda}$. Para os resultados apresentados nas Figuras 3.4(a) e 3.4(b) foi usado $\bar{\lambda} = 0,65$.

Na Figura 3.4(b), vemos que a distribuição de avalanches $P(S)$, é dada pela lei de potência, $P(S) = S^{-0,9\pm 0,1}$, indicando a existência de avalanches de todos os tamanhos.

Desde a sua proposição, a dinâmica **BS** tem sido amplamente estudada. Podemos ressaltar abordagem via campo médio [47–50] no qual os vizinhos são escolhidos aleatoriamente para a mutação ou evolução. Um outro trabalho que merece destaque é o artigo de Stefan Boettcher e Maya Paczuski [51], onde eles estimam os valores de adaptabilidade crítica para redes regulares de dimensão entre $1 \leq \mathbf{d} \leq 6$. Na Tabela 3.1 vemos os valores das adaptabilidade críticas, λ_c , para rede regulares de até 6 dimensões.

| \mathbf{d} | λ_c |
|--------------|-------------|
| 1 | 0,67 |
| 2 | 0,83 |
| 3 | 0,58 |
| 4 | 0,08 |
| 5 | 0,32 |
| 6 | 0,21 |

Tabela 3.1: Lista dos valores de adaptabilidade crítica, λ_c , para redes regulares de dimensões $\mathbf{d} = 1, 2, 3, 4, 5$ e 6 .

3.2.2 Otimização Extrema: simples e generalizada

O método *extremal optimization* (**EO**), que traduzimos como otimização extrema, foi criado por Stefan Boettcher e Allon G. Percus [10, 11], tem sua dinâmica inspirada na criticalidade auto-organizada, um conceito introduzido para descrever complexidade emergente em sistemas físicos, onde uma estrutura com adaptação otimizada emerge naturalmente, pela simples seleção contra o extremamente ruim. O método **EO**, assim como o *simulated annealing* [52] e o algoritmo genético [53] (que foram abordados

na minha dissertação de mestrado [54]), são inspirados em observação de sistemas reais da natureza.

A dinâmica **EO** simples consiste em, estando o sistema numa determinada configuração, encontrar a cada passo temporal a espécie que possui o pior valor de adaptabilidade. Em seguida, essa espécie tem seu valor de adaptabilidade atualizado (o que podemos interpretar como mutação ou seleção natural atuando sobre as espécies). Vale lembrar que os novos valores das variáveis, atualizadas, não são necessariamente melhores que os anteriores.

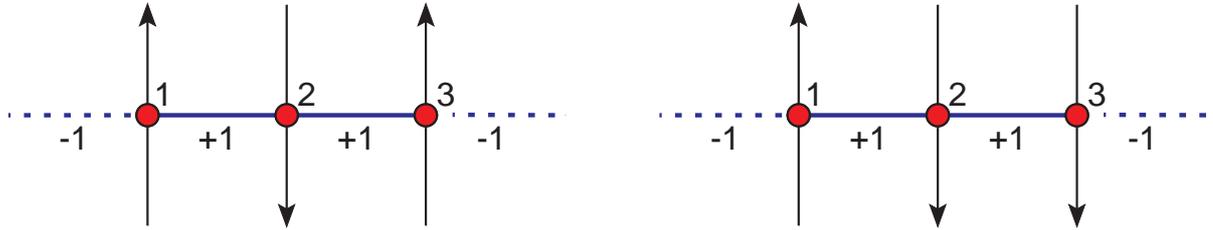
Diferente do que acontece com a dinâmica co-evolutiva **BS** onde as espécies com os piores valores de adaptabilidade e seus primeiros vizinhos tem seus valores de adaptabilidade atualizados, na dinâmica **EO** apenas a espécie com o pior valor de adaptabilidade é selecionado para mutação (extinção). Ou seja, na dinâmica **EO** não há interação entre a espécie selecionada para mutação e a as espécies vizinhas.

Para poder exemplificar o método **EO** simples, tomemos o sistema vidro de *spin* (*spin glass*). Trataremos aqui, do caso unidimensional, onde o *spin* no sítio i é conectado a cada um de seus vizinhos mais próximos j por meio de uma variável de ligação $J_{ij} \in \{-1, 1\}$ determinada aleatoriamente. O espaço de configurações Ω consiste de todas as configurações $S = (x_1, \dots, x_n) \in \Omega$ onde $|\Omega| = 2^n$.

O objetivo aqui é minimizar a função custo, ou hamiltoniana

$$C(S) = H(x) = -\frac{1}{2} \sum_{\langle ij \rangle} J_{ij} x_i x_j. \quad (3.1)$$

Devido ao efeito de frustração, que faz com que o sistema fique armadilhado em configurações de mínimos locais, é difícil alcançar a configuração do estado fundamental. Um exemplo de armadilha é mostrado na Figura 3.5. Essa dificuldade aumenta ainda mais com a dimensão do sistema.



(a) Fazendo uso da Equação 3.2, calculamos os valores de adaptabilidade λ_i do sistema e encontramos $\lambda_1 = -1$, $\lambda_2 = -1$ e $\lambda_3 = 0$.

(b) Depois de modificar o estado do terceiro *spin* do sistema, os valores de adaptabilidade λ_i do sistema são os seguintes: $\lambda_1 = 0$, $\lambda_2 = 0$ e $\lambda_3 = 1$.

Figura 3.5: Exemplo de uma configuração do sistema vidro de *spin* que está armadilhada. Segundo o algoritmo **EO** o terceiro *spin* da rede sempre vai ser escolhido para alterar seu estado, fazendo com que o sistema fique preso uma configuração de mínimo local.

Para encontrar a configuração de menor energia, o método **EO** simples atribui um valor λ_i (adaptabilidade ou qualidade) para cada variável de *spin* x_i

$$\lambda_i = x_i \left(\frac{1}{2} \sum_j J_{ij} x_j \right), \quad (3.2)$$

fazendo uso da equação (3.2) e reescrevendo a equação (3.1), temos que

$$C(S) = - \sum_{i=1}^n \lambda_i. \quad (3.3)$$

Assim como na dinâmica co-evolutiva de **BS**, apresentada na seção 3.2.1, o método **EO** simples evolui no espaço de configurações mudando o estado da variável que possui o pior valor de adaptabilidade em cada atualização (passo). A vizinhança $N(S)$ para uma atualização consiste de todas as configurações $S' \in N(S)$ que podem ser obtidas de S através da inversão de um único *spin*. Depois de cada atualização, a contribuição da variável que foi mudada e dos seus vizinhos serão recalculados pela equação (3.2).

O algoritmo do método EO simples tem os seguintes passos:

1. Iniciar uma configuração $S = \{x_1, \dots, x_n\} \in \Omega$, onde x_n é a variável dinâmica do sistema, e fazer $S_{\text{ótima}} := S$, onde $S_{\text{ótima}}$ é a melhor configuração a cada passo temporal;
2. Para a configuração atual;
 - a) Calcular λ_i para cada sítio i , onde λ_i é o valor de adaptabilidade do sítio i ;
 - b) Encontrar j que satisfaça $\lambda_j \leq \lambda_i$ para todo i , isto é, encontrar o sítio j de pior valor de adaptabilidade λ_j ;
 - c) Escolher aleatoriamente uma configuração $S' \in N(S)$, onde $N(S)$ é o espaço de configuração do sistema, requerendo que a pior variável x_j tenha seu estado alterado;
 - d) Aceitar $S := S'$ (incondicionalmente);
 - e) Se $C(S) < C(S_{\text{ótima}})$, onde $C(S)$ é a função a ser otimizada, então fazer $S_{\text{ótima}} := S$, caso contrário manter a configuração atual e procurar pelo novo pior valor de λ_i na configuração S ;
3. Repetir o passo 2 o quanto for necessário (teste de convergência).

O método **EO** se mostra bastante eficiente na busca de soluções de problemas de otimização [10, 11], e essa eficiência acontece quando a variável dinâmica assume vários estados. Esse não é o caso do sistema vidro de *spin*, que possui uma variável dinâmica capaz de assumir apenas dois estados (*spin up* ou *spin down*). A consequência disso é que as chances do sistema ficar “preso” numa configuração que não minimize a energia do sistema é muito grande, e nesse caso a dinâmica do método **EO** precisa sofrer uma modificação.

Para evitar que a dinâmica do método **EO** fique armadilhada (presa numa configuração de mínimo local), Boettcher e Percus introduziram um parâmetro τ (onde τ é um número real positivo), criando o algoritmo chamado de Otimização Extrema Generalizada **GEO** (do inglês, Generalized Extremal Optimization), também conhecido como

EO- τ [10, 11]. Antes de usar o parâmetro τ , é necessário fazer a ordenação (*rank*) de todas as variáveis x_i em uma lista de acordo com o valor de adaptabilidade λ_i , ou seja, é preciso encontrar uma permutação Π das variáveis de rótulo i com

$$\lambda_{\Pi(1)} \leq \lambda_{\Pi(2)} \leq \dots \lambda_{\Pi(n)}, \quad (3.4)$$

de tal modo que, no final do ordenamento teremos a pior variável x_j na posição 1 da lista, $j = \Pi(1)$, e a melhor variável na posição n da lista, onde n é número de vértices total da rede. Pressupondo que a distribuição de probabilidade obedeça a uma lei de potência

$$P_k \propto k^{-\tau}, \quad 1 \leq k \leq n, \quad (3.5)$$

teremos uma dinâmica para um dado valor do parâmetro τ . A cada atualização, é selecionada uma posição k na lista de acordo com o P_k . Então, escolhemos uma nova configuração (S'), de modo que a variável x_i com $i = \Pi(k)$ mude seu estado.

Com o algoritmo podendo escolher uma configuração diferente daquela que levaria obrigatoriamente à uma diminuição do valor da Equação 3.3 (função custo), a dinâmica **GEO** consegue se livrar de configurações de mínimos locais (atratores).

O algoritmo do método GEO tem os seguintes passos:

1. Iniciar uma configuração $S = \{x_1, \dots, x_n\} \in \Omega$ onde $|\Omega| = 2^n$, atribuir um valor para τ , e fazer $S_{\text{ótima}} := S$;
2. Para a configuração atual;
 - a) Calcular λ_i para cada sítio i ;

- b) Classificar as variáveis x_i de acordo com o valor (ordem de grandeza) de λ_i , isto é, encontrar uma permutação Π das variáveis de rótulo i com

$$\lambda_{\Pi(1)} \leq \lambda_{\Pi(2)} \leq \dots \leq \lambda_{\Pi(n)}.$$
 - c) De posse das variáveis dinâmicas ordenadas, calcular a distribuição de probabilidade para cada variável de rótulo k .
 - e) Modificar a variável x_i indicada pela distribuição de probabilidade P_k .
 - f) Aceitar $S := S'$ (incondicionalmente);
 - g) Testar se $C(S) < C(S_{\text{ótima}})$, se a condição for verdadeira, então fazer $S_{\text{ótima}} := S$, caso contrário, manter a atual configuração e procurar o novo pior valor de x_i na configuração S ;
3. Repetir o passo 3 o quanto for necessário (teste de convergência);

Dentre os problemas onde o método **GEO** foi aplicado, podemos citar dois dos mais conhecidos: o problema do grafo bipartido [11] e o do caixeiro viajante [53].

3.3 Dinâmicas Bak-Sneppen e Otimização Extrema Generalizada

Como vimos na Seção 3.2.1, no modelo **BS** o valor de adaptabilidade representa a habilidade de sobrevivência de uma população inteira de uma espécie. Essa é uma quantidade vital e a única utilizada nessa dinâmica co-evolutiva. Na Seção 3.2.2, vimos o método de **GEO**, que foi criado por Stefan Boettcher e Allon G. Percus [10, 11], e que foi inspirado na criticalidade auto-organizada. Por levar o sistema estudado para (ou muito próximo do) o estado fundamental rapidamente, o método **GEO**, tem se mostrado muito eficiente [11].

Nas próximas seções deste capítulo, faremos uso da dinâmica **GEO** biológica (**GEO-b**), que é ligeiramente diferente da **GEO**. A diferença entre as duas dinâmicas é a seguinte:

- GEO** Como podemos ver da Equação 3.2, o valor de adaptabilidade λ_i é calculado a partir hamiltoniana do sistema;
- GEO-b** Para essa dinâmica, o valor de adaptabilidade λ_i é gerado de uma distribuição uniforme.

Compreendida a diferença entre os algoritmos, construímos o algoritmo da dinâmica **GEO-b**, da seguinte forma:

- 1) Para cada sítio i de uma rede com N sítios, é associado um valor de adaptabilidade λ_i com valores entre 0 e 1, escolhido aleatoriamente de uma distribuição uniforme;
- 2) Todos os vértices da rede são classificados em ordem crescente de acordo com os valores de suas adaptabilidades. Desta forma, o sítio com o menor valor de adaptabilidade é o primeiro da lista;
- 3) Um sítio de posição k na lista, ($1 \leq k \leq N$), é selecionado segundo uma probabilidade $P(k) \sim k^{-\tau}$, onde τ é um número real positivo definido, e o valor de adaptabilidade λ_i do sítio da posição k é alterado para λ'_i , gerado de uma distribuição uniforme;
- 4) Os passos 2) e 3) são repetidos tantas vezes quantas forem necessárias.

Na Seção 3.3.1 faremos um estudo comparativo da eficiência das dinâmicas **BS** e **GEO-b**. Utilizaremos como parâmetro de eficiência o valor de adaptabilidade média das espécies na rede (sistema). Uma dinâmica é considerada mais eficiente do que outra se conseguir elevar mais rápido o valor de adaptabilidade média do sistema. Esse valor é medido da seguinte forma: estando no estado estacionário, somamos todos os valores de adaptabilidade e dividimos pelo número de sítios da rede, fazemos isso a cada passo

temporal e armazenamos o resultado. No final da simulação somamos o resultado da adaptabilidade média de cada passo e dividimos pelo tempo de simulação efetivo, que é obtido subtraindo o tempo necessário para o sistema entrar no estado estacionário do tempo de total da simulação.

Ainda Seção 3.3.1, apresentaremos uma análise de uma forma discreta do modelo **BS**, proposto por nós e que chamamos de Bak-Sneppen Discreto **BSD**, afim de verificar se a biodiversidade de espécies é um requisito essencial para alcançar (ou manter) a criticalidade auto-organizada. A diferença entre as dinâmicas **BSD** e **BS** está na atribuição dos valores de adaptabilidades das espécies. Na dinâmica **BS** o valor de adaptabilidade, λ_i , de cada espécie é sorteado de uma distribuição uniformemente distribuída entre 0 e 1. No caso da dinâmica **BSD** o valor de adaptabilidade, λ_i , é atribuída fazendo $\lambda_i = q/Q$, onde Q (um número inteiro) representa o número de adaptabilidades possíveis e q é um número aleatório que pode assumir os valores 1, 2, ..., ou Q . De modo prático, podemos interpretar a discretização dos valores de adaptabilidade como sendo uma diminuição da biodiversidade do sistema. Devido a essa discretização, um número grande de espécies que possuem o mesmo valor de adaptabilidade podem coexistir. Com isso, surge a seguinte dúvida, uma vez que poderá existir várias espécies portadoras do menor valor de adaptabilidade, qual espécie deve ser escolhida para mutação? Uma solução para esse problema é colocar todas as espécies numa lista e sortear aleatoriamente uma delas.

Como veremos nas Seções 3.3.1 e 3.3.2, a dinâmica **BSD** pode não possuir a característica de criticalidade auto-organizada, dependendo do nível de discretização da sua variável de adaptabilidade. Assim como na natureza, a biodiversidade se mostra fundamental para os modelos evolutivos teóricos, sem a qual a criticalidade auto-organizada não emerge. Contudo, mesmo trabalhando com um conjunto finito de valores de adaptabilidade, a ausência de biodiversidade pode ser contornada incorporando um ruído η ao valor de adaptabilidade da dinâmica **BSD**, produzindo uma nova dinâmica, também pro-

posta por nós e que chamamos de Bak-Sneppen com Ruído **BSR**, que será apresentada em detalhes na Seção 3.3.2.

Seguindo o estudo da biodiversidade de espécies, analisamos o modelo **BSR** para descobrir a partir de que valor de ruído faz com que a biodiversidade seja suficiente para o surgimento das características de criticalidade auto-organizada.

3.3.1 Comparação entre as Dinâmicas **BS**, **BSD** e **GEO-b**

*P*ara comparar as dinâmicas **BS** e **GEO-b**, fizemos simulações com $1,1 \times 10^9$ passos temporais em um anel unidimensional com $N = 4\,001$ sítios. Para garantir que o regime do estado estacionário já tenha sido alcançado, descartamos as primeiras $1,0 \times 10^8$ rodadas, assim a média temporal é feita sobre o número de rodadas restantes. Na Figura 3.6 vemos que a frequência de adaptabilidade, da dinâmica **BSD** com $Q = 10$, parece não possuir um limiar crítico λ_c bem definido. O valor de adaptabilidade média da dinâmica **BSD** de 0,86 (veja Figura 3.7) é um pouco maior do que o **BS** tradicional.

Ainda na Figura 3.6, vemos claramente que para $\tau = 0,05$ o algoritmo **GEO-b** se comporta como uma caminhada aleatória e tem uma distribuição de adaptabilidade quase uniforme. Para $\tau = 0,5$ a distribuição é uma função linear crescente de λ , enquanto para $\tau = 1,0$ a distribuição cresce exponencialmente com λ . Para a dinâmica **BS**, a distribuição tem a forma de uma função escada com uma descontinuidade no ponto crítico $\lambda_c \sim 0,67$. Esse ponto crítico existe em todas as redes regulares ou em redes exponenciais (redes com distribuição de conectividades escalando com uma Poisson, como a rede Strogatz-Watts [31]), mas não existe em redes livres de escala. A existência desse ponto crítico no algoritmo **BS** é o primeiro sinal de que há um estado crítico auto-organizado. A dinâmica **GEO-b** não possui esse ponto crítico, e tampouco possui criticalidade auto-organizada.

Para medir a eficiência dos algoritmos **BS**, **BSD** e **GEO-b**, mostramos na Figura 3.7 a adaptabilidade média obtida no regime de estado estacionário. Para a dinâmica **BS** a adaptabilidade média é 0,83. Essa adaptabilidade média corresponde a dinâmica **GEO-b** com $\tau = 0,86$. Para $\tau = 1,5$ a adaptabilidade média da dinâmica **GEO-b** é aproximadamente 0,99. Isso significa que para valores altos de τ , o algoritmo **GEO-b** cria um meio ambiente ideal, onde apenas as espécies mais aptas sobrevivem. Para τ no intervalo $[0, 0,86]$ a dinâmica **BS** supera o **GEO-b**.

Ainda, comparando as dinâmicas co-evolucionária **BS**, **BSD** e evolucionária **GEO-b**, analisamos a correlação espacial $C(x)$. $C(x)$ é a distribuição de probabilidade de

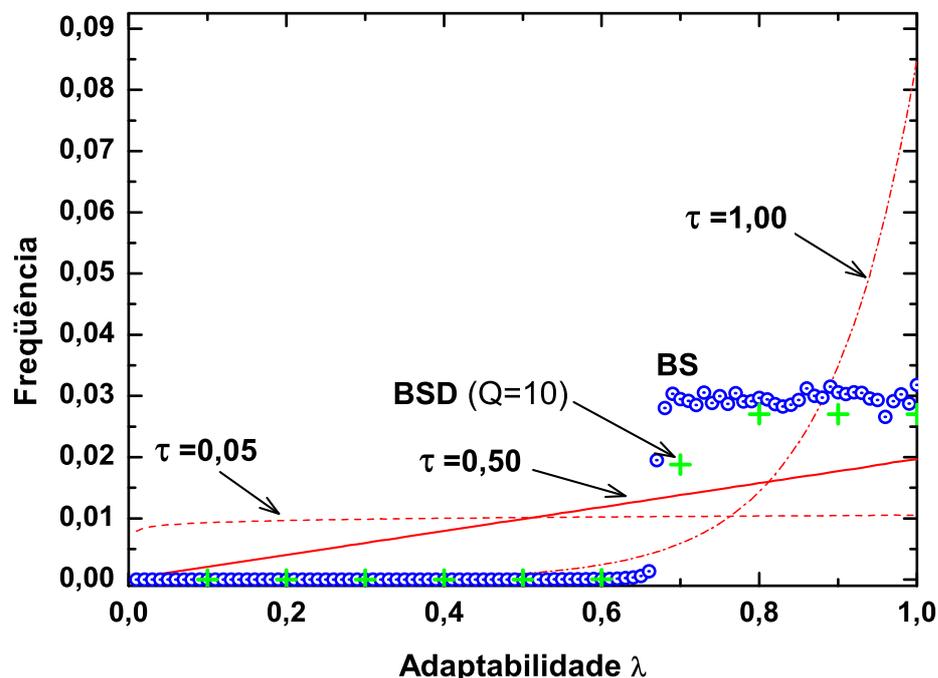


Figura 3.6: Frequência de adaptabilidade λ medida depois de alcançado o regime estacionário. Os círculos azuis correspondem ao algoritmo **BS** e exibe uma descontinuidade no ponto crítico $\lambda_c \sim 0,67$. As cruces verdes representam o algoritmo **BSD** com apenas 10 valores de adaptabilidade possíveis. As linhas tracejada, contínua e tracejada com pontos correspondem ao algoritmo **GEO-b** para os valores de $\tau = 0,05$, $0,50$ e $1,00$, respectivamente.

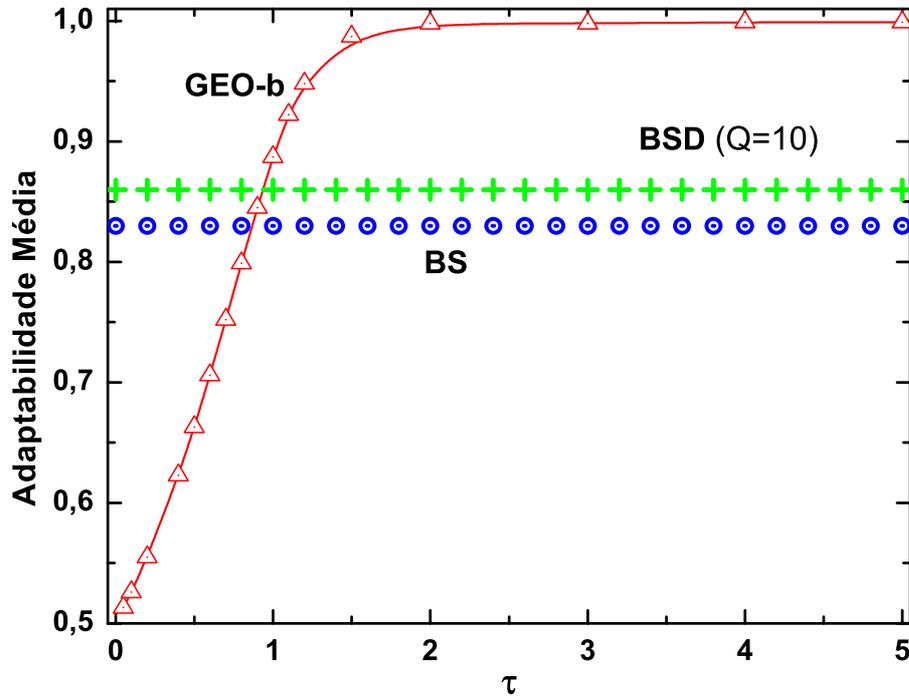


Figura 3.7: Adaptabilidade média versus τ . O algoritmo **BS** tem um valor de adaptabilidade média de 0,83 que corresponde ao valor de $\tau = 0,86$ da dinâmica **GEO-b**, enquanto o valor de adaptabilidade média para o algoritmo Bak-Sneppen Discreto **BSD** é 0,86 e corresponde a $\tau = 0,95$ da curva da dinâmica **GEO-b**.

distância x entre duas espécies extintas (ou que sofreram mutações) subsequentes. Apresentamos na Figura 3.8 os resultados dessa comparação. Vemos que a dinâmica **GEO-b** não apresenta um comportamento crítico auto-organizado na distribuição de probabilidade de distância, pois ao invés de uma lei de potência para $C(x)$, a dinâmica **GEO-b** possui uma correlação espacial com um alcance infinito — a distribuição de probabilidade $C(x)$ é constante não importando qual seja distância entre duas espécies subsequentemente extintas ou modificadas. No caso da dinâmica **BS**, vemos que a distribuição de probabilidade $C(x)$ obedece a lei de potência, $C(x) \sim x^{-3,23 \pm 0,02}$ — um indício de criticidade auto-organizada.

Outra importante diferença entre as duas dinâmicas é a distribuição de probabilidade de avalanches $P(A)$, onde A é o tamanho de uma avalanche e é definido pelo número de passos temporais subseqüentes com pelo menos uma espécie com valor de adaptabilidade abaixo do limiar crítico λ_c . Como podemos ver na Figura 3.6, esse limiar crítico não existe na dinâmica **GEO-b**. Para o algoritmo **BS** a distribuição de avalanches decai como $P(A) \sim A^{-0,99 \pm 0,01}$ para um valor de adaptabilidade $\lambda_c = 0,67$, veja Figura 3.9(a). Enquanto no algoritmo **GEO-b**, a distribuição de avalanches decai segundo uma lei de potência truncada com $A_c(\lambda_c, \tau)$ (veja Figura 3.9(b)), e depende dos valores escolhidos de λ_c e τ .

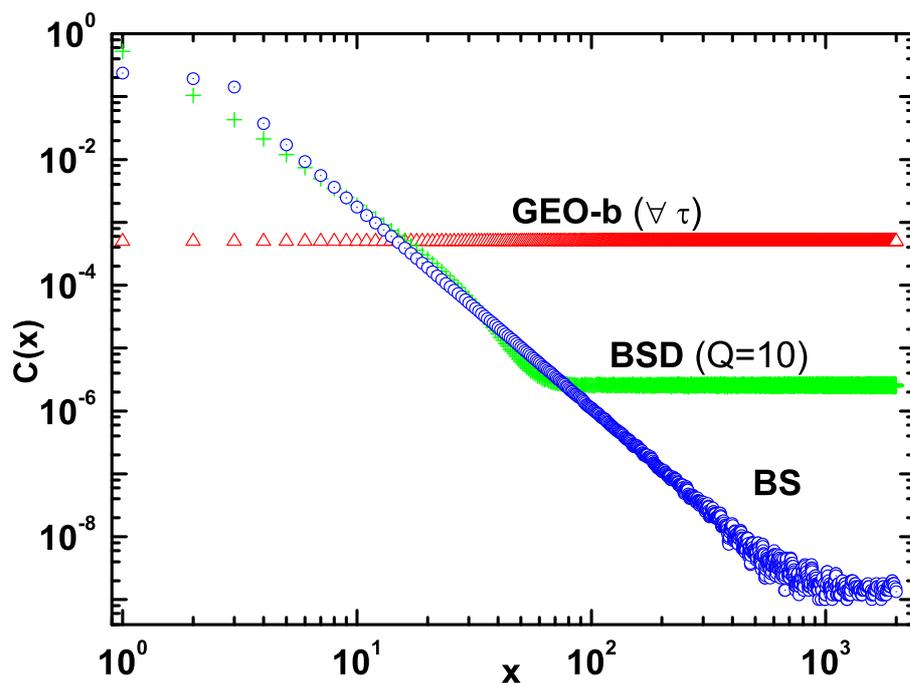
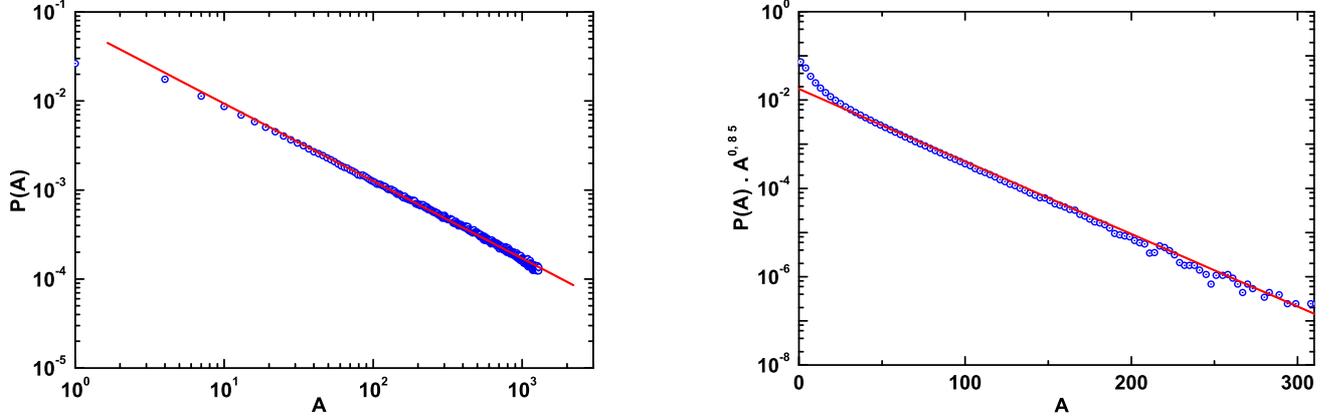


Figura 3.8: A dinâmica **GEO-b** tem uma distribuição de probabilidade $C(x)$ constante que independe do valor de τ . O algoritmo **BSD** tem um comportamento misto enquanto o algoritmo **BS** segue uma lei de potência.



(a) Distribuição de probabilidade de avalanches $P(A)$, da dinâmica **BS**, escalando com $P(A) \sim A^{-0,99 \pm 0,01}$.

(b) Distribuição de probabilidade de avalanches $P(A)$, da dinâmica **GEO-b**, com valores de $\lambda_c = 0,85$ e $\tau = 1$, que apresenta um comportamento misto dado por $P(A) \sim A^{-0,85} e^{-0,02A}$.

Figura 3.9: Distribuição de probabilidade de avalanches das dinâmicas **BS** e **GEO-b**.

3.3.2 Dinâmicas BSD e BSR

Na Seção 3.3.1 vimos que, a dinâmica **BSD** não apresenta a característica de criticalidade auto-organizada. Assim como na natureza, a biodiversidade parecer ter um papel fundamental em modelos evolutivos, uma vez que sem ela a criticalidade auto-organizada não emerge. Para descobrir o momento em que acontece a transição da dinâmica **BSD** para a **BS**, incorporamos um ruído η ao valor de adaptabilidade da dinâmica **BSD**,

$$\lambda = q/Q + (1 - 2r)\eta, \quad (3.6)$$

onde Q (um número inteiro) representa o número de adaptabilidades possíveis, q é um número aleatório que pode assumir os valores $1, 2, \dots$, ou Q , r é um número aleatório gerado por uma distribuição uniforme no intervalo $[0, 1]$ e η é um ruído (um número real, muito pequeno) que vai de 10^{-3} e 10^{-8} .

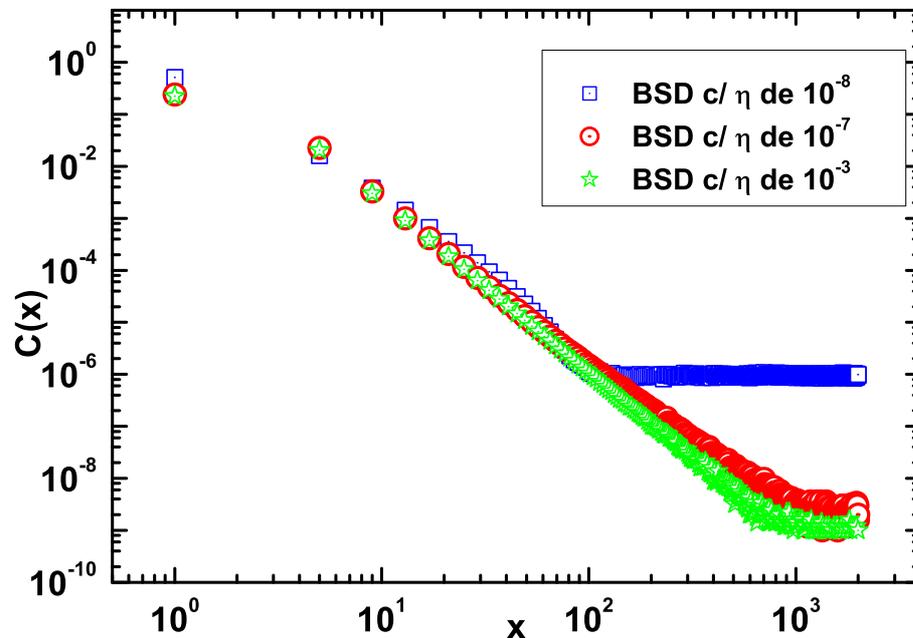


Figura 3.10: Distribuições de probabilidades $C(x)$ para a dinâmica **BSD** com $Q = 10$, para valores de $\eta = 10^{-3}$, 10^{-7} e 10^{-8} .

Como podemos ver na Figura 3.10, para valores de η maiores do que 10^{-7} a dinâmica **BSR** possui a característica de criticalidade auto-organizada. Esse ruído pode ser interpretado como o surgimento de sub-espécies, contribuindo para o restabelecimento da biodiversidade do sistema.

Nessa mesma linha, vemos na Figura 3.11, uma comparação entre as dinâmicas **GEO-b**, **BSD** e **BSR**. A curva da distribuição de probabilidade espacial $C(x)$ da dinâmica **BSR**, com um valor de ruído (η) de 10^{-3} , recupera completamente a característica de criticalidade auto-organizada.

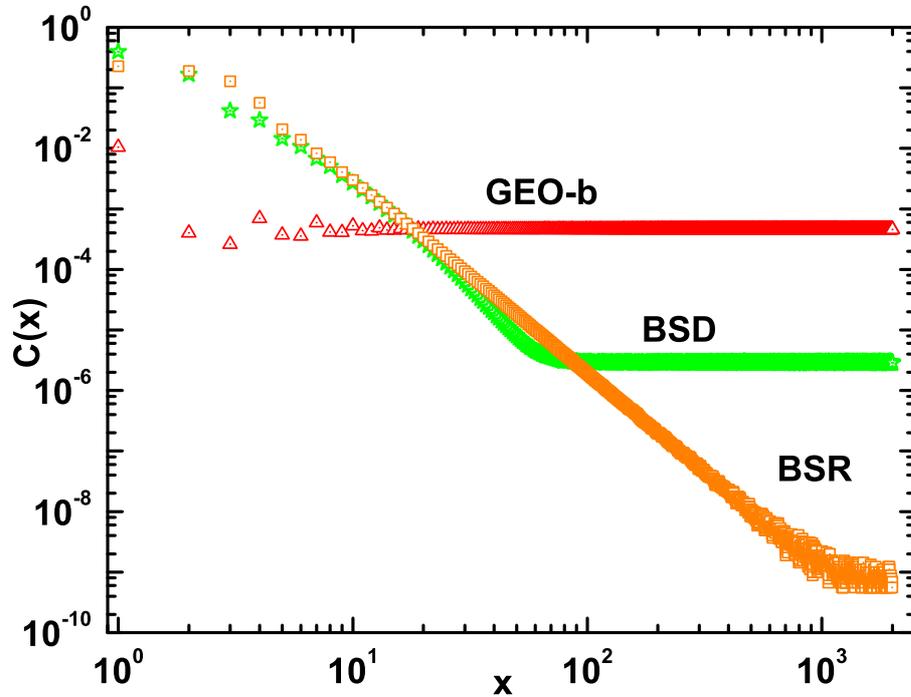


Figura 3.11: Como já vimos na Figura 3.9 o algoritmo **GEO-b** tem uma distribuição de probabilidade $C(x)$ constante que independe do valor de τ . O algoritmo **BSD** tem um comportamento misto enquanto o algoritmo **BSR**, com um ruído (η) de apenas 10^{-3} , recupera o decaimento tipo lei de potência.

3.4 Resultados

Pelo fato de levar as espécies menos adaptadas a extinção, a teoria co-evolucionária de Darwin é intrinsecamente uma teoria de otimização. Nesse capítulo estudamos a eficiência de alguns algoritmos. Comparando os valores médios de adaptabilidade dos algoritmos **BS** e **GEO-b**, concluímos que a dinâmica **BS** é mais eficiente do que a otimização extrema no intervalo $[0, 0,86]$ de τ . A determinação da correlação espacial e a distribuição de probabilidade de avalanches mostrou que a dinâmica de otimização extrema não conduz o sistema para um estado crítico auto-organizado. Por meio de uma

forma discreta do modelo Bak-Sneppen, que chamamos de modelo Bak-Sneppen Discreto **BSD**, chegamos a conclusão de que a biodiversidade é um requisito essencial para o surgimento da criticalidade auto-organizada.

Como vimos no Capítulo 2, enquanto **BS** é uma dinâmica co-evolucionária a dinâmica **GEO-b** é apenas evolucionária, ou seja, as espécies sujeitas a essa dinâmica não interagem. Comparações do desempenho de dinâmicas co-evolucionárias e evolucionárias foram feitas no contexto de autômato celular [55].

Observamos ainda que, a dinâmica do algoritmo **BS** é uma descrição, simplificada, da evolução biológica adotada pela natureza enquanto a dinâmica do algoritmo **GEO-b** representa uma otimização criada pelo homem. Ao contrário do que acontece com a dinâmica **BS**, a dinâmica **GEO-b** não tem a característica co-evolucionária, ou seja, a extinção de uma espécie não influencia no destino de espécies vizinhas. A dinâmica **GEO-b** também não tem a característica de eliminação Darwiniana, uma vez que a espécie eliminada não é necessariamente a menos adaptada.

Apesar da eficiência do algoritmo **GEO-b** poder exceder, em certos casos (se $\tau > 0,86$) o algoritmo **BS**. Essa eficiência é acompanhada das seguintes características indesejáveis: a correlação espacial entre as espécies é constante, ou seja, ela independente da distância entre espécies subsequentemente extintas; existe um parâmetro externo livre τ para ser ajustado manualmente; inexistência de um mecanismo de equilíbrio entrecortado. Tal mecanismo parece ser uma boa maneira encontrada pela natureza de renovar, por mutação, espécies sem a intervenção de mudanças climáticas bruscas ou destruição em massa (por meteoro, por exemplo). Aprendemos que a dinâmica **GEO-b** não conduz o sistema a um estado crítico auto-organizado.

No capítulo 4, apresentaremos uma breve revisão histórica e de conceitos fundamentais para a teoria de grafos ou redes complexas.

CAPÍTULO 4

Redes, O Que São E Como Surgiram

Tradicionalmente os cientistas buscam modelos mínimos. Como veremos na breve revisão histórica feita neste capítulo, modelos mínimos são a “pedra filosofal” tanto dos físicos como dos matemáticos, uma vez que esses profissionais das ciências exatas estão sempre a procura de modelos mínimos para descrever seus sistemas. Em nossa breve revisão apresentaremos um pouco da história do nascimento da teoria de grafos, com Euler e as sete pontes de *Konigsberg* [56, 57], e seguiremos apresentando os modelos de Erdős-Rényi [58–60], de Watts-Strogatz [31, 61, 62] e de Barabási-Albert [63–66], chegando aos dias de hoje com as várias aplicações das redes no estudo de sistemas reais.

4.1 Um pouco da vida de Euler e as Pontes de *Konigsberg*

Teorema 1. Seja $M_\omega = \frac{n}{2}(\log n + \omega)$, onde $\omega = \omega(n)$ é uma função de n . Se $\omega \rightarrow -\infty$ então um G_{n, M_ω} característico é desconectado, enquanto se $\omega \rightarrow \infty$ um G_{n, M_ω} característico é conectado.

Não, a teoria de grafos e conseqüentemente a ciência das redes não começaram com teoremas quase que incompreensíveis. Mas também não começaram com definições de fácil compreensão, assim como: consideremos uma rede, como sendo um

conjunto de ítems que podemos chamar de vértices ou nós, que estão conectados segundo alguma regra de ligação (veja a Figura 4.1).

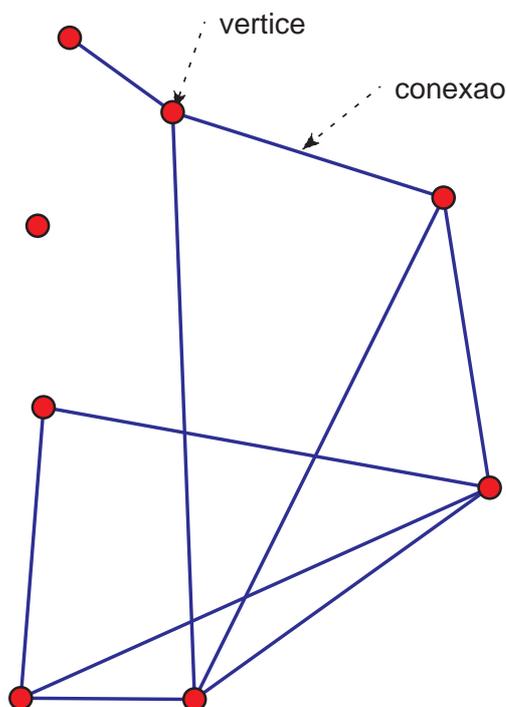


Figura 4.1: Pequena rede com oito vértices e dez conexões.

Tudo começou com Leonhard Euler¹, um matemático e físico de origem suíça. Seu pai, o pastor calvinista Paul Euler, desprezando o prodigioso talento matemático do filho determinou que ele estudasse teologia e seguisse a carreira religiosa. Daniel e Nikolaus Bernoulli, junto com Johann Bernoulli (um de seus futuros professores), convenceram Paul a permitir que o filho trocasse o hábito pelos números. Com certeza áreas como a matemática, a física e a engenharia têm muito a agradecer a esse homem. Euler redigia trabalhos com tanta facilidade que costumava dizer ironicamente que seu lápis era muito mais inteligente que ele. Suas obras completas estendem-se por dezenas de volumes, alguns dos quais infelizmente nunca foram finalizados². A gratidão não é somente

¹Euler nasceu na Basileia em 15 de Abril de 1707 e faleceu em 18 de Setembro de 1783, de ataque cardíaco

²Como a *Opera Omnia*, uma coletânea dos trabalhos de Euler (infelizmente incompleta), de mais de setenta e três volumes, cada volume contendo da ordem de seiscentas páginas.

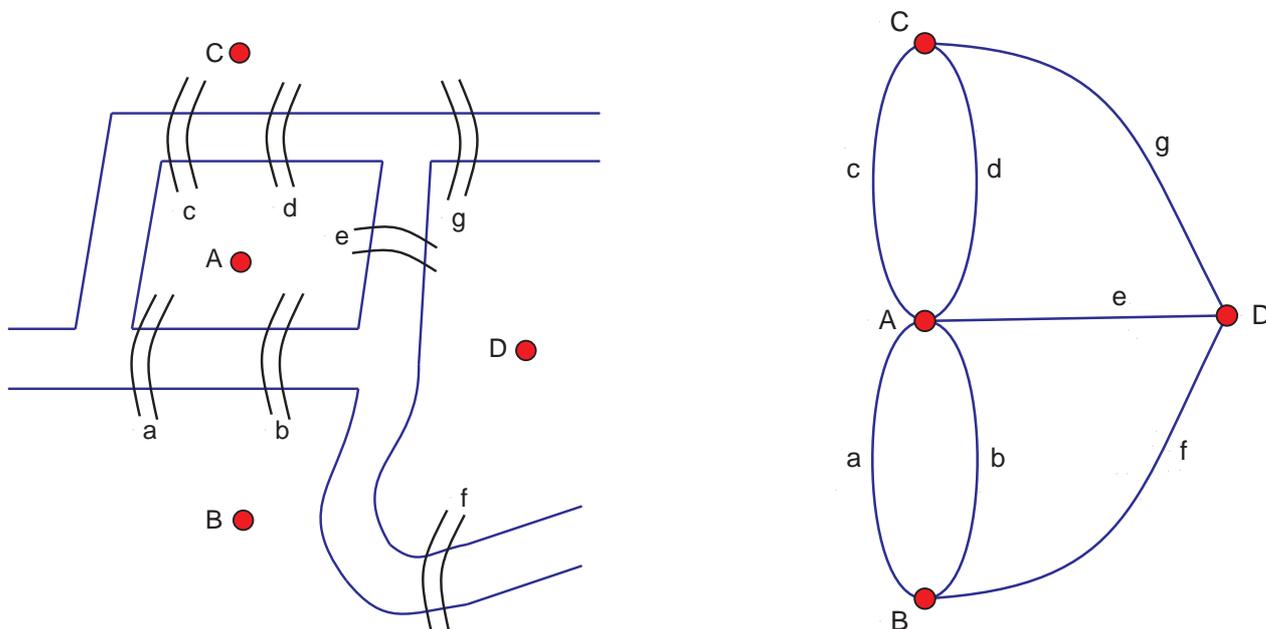
por suas grandes obras no campo da matemática pura, mas também pelos vários “desvios de percursos” que ele acabou fazendo ao longo de sua vida. Além da teoria dos números, Euler teve grande participação no desenvolvimento do cálculo diferencial e integral, na teoria das equações diferenciais, na geometria e muitas áreas da matemática aplicada; sendo um dos pioneiros também da teoria de grafos e da topologia algébrica.

Os últimos 10 anos da vida de Euler, entre seu retorno a *St. Petersburg* em 1766 e sua morte aos 76 anos, foram de muito tumulto. Apesar das várias tragédias pessoais, metade ou mais de seus trabalhos foi escrito justamente durante esse turbulento período de sua vida. Dentre esses trabalhos está um tratado de 775 páginas sobre o movimento da Lua, um livro texto de álgebra e um livro (composto de três volumes) sobre cálculo integral, que foi completado enquanto ele publicava uma média de um artigo sobre matemática por semana no jornal da Academia de *St. Petersburg*. No ano de 1766 Euler sofreu com a perda parcial de sua visão e no ano seguinte após uma fracassada operação de catarata Euler, perdeu totalmente a visão. Mas isso não o perturbou em nada e freqüentemente dizia: “agora eu terei menos distrações”. Devido a perda da visão, nesse último período de sua vida Euler não pôde ler e nem escrever uma única linha. Assim, milhares de páginas de trabalho tiveram que ser ditados a um ajudante que as transcrevia.

Três décadas antes, ainda com sua visão perfeita, Euler escreveu um curto artigo sobre um divertido problema que teve como palco *Konigsberg*³, uma cidade não muito distante da casa de Euler em *St. Petersburg*. Devido a existência do rio Pregel e da vocação da cidade para os negócios, uma vez que possuía uma frota muito requisitada de navios, os mercadores locais e seus familiares podiam desfrutar de uma vida confortável. As boas condições econômicas da cidade eram tais que os responsáveis pela mesma (prefeitos) construíram nada menos do que sete pontes em vários pontos da cidade. A maioria das pontes, 5 de um total de 7, conectavam a elegante ilha Kneiphof, cercada pelo rio Pregel, com o restante da cidade. As outras duas pontes ligavam outros pontos da cidade,

³*Konigsberg*, atualmente conhecida como Kaliningrado, uma cidade a leste da Prússia.

veja a Figura 4.2(a). A população de *Konigsberg*, aproveitando o clima de harmonia e a paz da cidade, no tempo livre se divertia com o seguinte quebra-cabeças: é possível encontrar um caminho atravessando as sete pontes sem nunca atravessar uma mesma ponte duas vezes? Contudo, ninguém foi capaz de encontrar um caminho que satisfizesse o quebra-cabeças até que uma nova ponte fosse construída em 1 875. Entretanto, quase 150 anos antes da construção da nova ponte, em 1 736, Euler ofereceu uma rigorosa prova matemática dizendo que não existia um caminho que passasse por todas as sete pontes uma única vez. Além de resolver o problema das pontes de *Konigsberg* (veja a Figura 4.2), Euler, de forma não intencional, iniciou uma nova área da matemática conhecida como teoria de grafos. Essa teoria é hoje a base de todo nosso conhecimento sobre redes. Durante os séculos que se seguiram a teoria de grafos cresceu e se desenvolveu com a contribuição de vários outros matemáticos.



(a) Pontes de *Konigsberg*. Um esboço fiel de *Konigsberg* antes de 1 875, com a ilha Kneiphof (A) e as áreas de terra (B, C e D) envolvidas pelos dois braços do rio Pregel.

(b) Grafo das 7 pontes de *Konigsberg*

Figura 4.2: Figura 4.2(a) mostra um esboço das pontes de *Konigsberg* e o grafo correspondente é mostrado na Figura 4.2(b).

4.2 Nascimento da Ciência das Redes

Como mencionado na seção 4.1, o grafo das pontes de *Konigsberg* foi o marco inicial para a teoria de grafos. Euler provou de modo simples e elegante que não é possível fazer uma caminhada pela cidade e passar por cada uma das 7 pontes uma única vez. A prova de Euler é compreensível mesmo para uma pessoa que não tenha intimidade com a matemática. No entanto, não foi a prova que entrou para a história, mas sim os passos que ele usou para resolver o problema. A grande idéia foi, sem dúvida alguma, tratar o problema das pontes de *Konigsberg* como sendo um grafo, uma coleção de vértices conectados por ligações. Usando vértices para representar cada uma das quatro áreas de terra separadas pelo rio, associando a cada uma, as letras **A**, **B**, **C** e **D**. O próximo passo foi usar ligações para representar cada uma das sete pontes, associando cada uma das pontes, às letras **a**, **b**, **c**, **d**, **e**, **f** e **g**.

A prova de Euler, de que em *Konigsberg* não existia um caminho tal que fosse possível passar por todas as sete pontes uma única vez, foi baseada na simples observação de que vértices com um número ímpar de ligações devem ser escolhidos para se começar ou terminar a caminhada, de forma que um caminho contínuo que passe por todas as pontes pode ter apenas um ponto de partida e um outro ponto de chegada. Assim, tal caminho não pode existir em um grafo que tem mais de dois vértices com um número ímpar de ligações. Visto que o grafo de *Konigsberg* tem os seus quatro vértices nessas condições, um caminho passando uma única vez por todas as pontes não existe.

O aspecto mais importante da prova de Euler é que a existência de possíveis caminhos não depende de quão inteligente somos para encontrá-los. A existência do caminho é uma propriedade intrínseca do grafo, ou seja, ou o caminho existe ou não existe. Assim sendo, dado o grafo das pontes de *Konigsberg*, nunca teremos sucesso na busca de um caminho que passe uma única vez por cada uma das sete pontes.

No ano de 1875, foi construída em *Konigsberg* uma nova ponte entre as áreas de terra **B** e **C** aumentando assim o número de ligações destes dois vértices para quatro. Com isso apenas os vértices **A** e **D** estavam com um número ímpar de ligações, sendo então possível encontrar o caminho desejado. A mensagem não intencional de Euler é muito simples: grafos (ou redes) têm propriedades escondidas em suas construções, que limitam ou aumentam nossa habilidade para fazer algo útil com elas.

Depois de Euler, a teoria de grafos continuou crescendo com as contribuições de grandes matemáticos tais como Cauchy⁴, Hamilton⁵, Cayley⁶, Kirchhoff⁷ e Polya⁸. Muito do que é conhecido hoje sobre os grafos, por isso podemos dizer seguramente que eles tiveram um papel decisivo para consolidar a teoria grafos. Até a metade do século XX, um dos objetivos em teoria de grafos era propor novos modelos e catalogar as suas propriedades.

Afinal de contas, como uma rede real se forma? Quais são as leis que governam o surgimento de cada um de seus elementos e estrutura? Estas questões, e a primeira resposta, não vieram antes da década de 60 do século XX, quando dois matemáticos húngaros, Paul Erdős e Alfréd Rényi revolucionaram a teoria de grafos.

Paul Erdős escreveu mais de 1500 artigos científicos até sua morte em 1996, sendo 8 desses artigos com Alfréd Rényi. Esses oito artigos pela primeira vez abordavam a questão mais fundamental para o entendimento do mundo que em que vivemos: Como as redes se formam?

⁴Augustin-Louis Cauchy (21/08/1789, Paris - 23/05/1857, Sceaux (próximo a Paris)) foi um matemático francês e contribuiu muito no estudo do cálculo de resíduos.

⁵William Rowan Hamilton (04/08/1805, Dublin - 02/09/1865, Dublin). Hamilton é o maior homem de ciência que a Irlanda produziu. Uma de suas heranças é álgebra dos quatérnios publicado no livro "Lectures on Quaternions".

⁶Arthur Cayley (16/08/1821, Richmond, Inglaterra - 26/01/1895, Cambridge, Inglaterra) foi um matemático britânico. As suas contribuições incluem a multiplicação de matrizes e o teorema de Cayley.

⁷Gustav Robert Kirchhoff (12/03/1824 em Königsberg, Prússia - 17/10/1887 em Berlim, Alemanha) foi um físico alemão cujas principais contribuições científicas estiveram no campo dos circuitos elétricos, na espectroscopia e na emissão de radiação dos corpos negros. Kirchhoff propôs o nome de "radiação do corpo negro" em 1862. É o responsável por duas leis fundamentais da teoria clássica dos circuitos elétricos e da emissão térmica.

⁸George Polya (13/12/1887, Budapeste - 07/09/1985, Palo Alto) foi um matemático húngaro. Dentre seus trabalhos mais conhecidos podemos citar a solução do passeio aleatório.

Para responder a essa pergunta, durante as conversas de trabalho, Erdős e Rényi, propuseram o seguinte experimento mental: suponhamos que uma festa seja organizada, e que sejam convidadas 100 pessoas com a seguinte condição, ninguém se conhece. Ofereça a este grupo de estranhos vinho e queijo, logo eles estarão se organizando em grupos, formados por duas, três ou mais pessoas, e estarão conversando. Agora mencione a um dos convidados que o vinho tinto nas garrafas verdes é um saboroso e raro vinho tinto, vintage do Porto, de vinte anos de idade, muito melhor do que o vinho tinto das outras garrafas. Mas peça ao seu convidado para somente compartilhar essa informação com os novos amigos que ele fez na festa. Pois a quantidade de vinho do Porto é pequena e não seria o suficiente para todos. Contudo, os convidados, cansados de conversar muito tempo com as mesmas pessoas, irão iniciar uma conversa com outras pessoas, entrando em novos grupos de conversa. Um dos assuntos mais interessantes nesses novos grupos será a existência do raro vinho do Porto, mesmo que ninguém note nada de diferente. Passará a existir uma rede social, formadas pelas ligações entre pessoas que se reuniram em grupos e ligações entre grupos, formada pelas ligações feitas por aqueles que mudaram de grupo. Conseqüentemente, poderíamos identificar os caminhos ligando pessoas até então desconhecidas umas das outras.

Os convidados da festa podem ser pensados como parte de um problema de teoria de grafos, onde eles são os vértices e uma conversa cria uma ligação social entre eles. Assim, desta rede de relacionamentos — um grafo emerge — com vértices conectados por ligações. Computadores ligados por linhas telefônicas, moléculas ligadas por reações químicas, fábricas e consumidores ligados por mercadorias, células nervosas conectadas por axônios, ilhas ligadas por pontes, são todos exemplos de grafos. Erdős e Rényi, começaram a buscar um modelo que pudesse descrever todos estes diferentes sistemas, um verdadeiro desafio para qualquer um. Indiferentes a dificuldade, e tendo sempre em mente que o objetivo maior de um cientista é encontrar a maneira mais simples de explicar um fenômeno complexo, a resposta deles a esse desafio não poderia ser mais simples.

Uma vez que diferentes sistemas seguem regras muito diferentes para a formação de suas redes, Erdős e Rényi, diante da diversidade dos sistemas, propuseram que as ligações entre os vértices fossem aleatórias. Dado um conjunto de n vértices, conecte os vértices dois a dois com uma probabilidade p , onde p é um número real entre 0 e 1.

Erdős se questionava: “Criamos a matemática ou apenas a descobrimos?”. Podemos estender o alcance dessa pergunta, reformulando-a da seguinte forma: “As verdades existem, mesmo que as desconheçamos?”. Erdős tinha uma resposta para esta questão, segundo ele: “As verdades matemáticas estão entre uma lista de verdades absolutas, e apenas as re-descobrimos”.

Em oposição ao seu amigo Albert Einstein, em Princeton, que dizia “Deus não joga dados com o universo”, Erdős se aventurava dizer que “Deus se divertia jogando dados”.



4.3 O Mundo é Pequeno

Em 1967, Stanley Milgram, um psicólogo social e professor em Harvard, se lançou em mais um de seus experimentos sociais. O objetivo de Milgram agora era encontrar a “distância” entre duas pessoas quaisquer nos Estados Unidos da América [67]. Para dar início a esse experimento duas pessoas foram escolhidas como alvo. A primeira foi a esposa de um estudante de teologia em Sharon, Massachusetts. A segunda foi um corretor em Boston. O próximo passo foi escolher as cidades de Wichita, Kansas e Omaha, todas elas no Nebraska, como ponto de partida para o estudo.

Havia pouco consenso sobre quantas ligações seriam necessárias para conectar duas pessoas dos estados de Massachusetts e Nebraska ou dos estados de Boston

e Nebraska. Tanto que o próprio Milgram, enquanto planejava o experimento, disse: “Recentemente eu perguntei a um amigo muito culto quantos seriam os passos para ligar duas pessoas, e ele respondeu que acreditava serem necessários 100 passos (pessoas), ou mais, para fazer uma carta chegar de Nebraska até Sharon”.

O experimento de Milgram exigiu o envio de cartas para moradores, aleatoriamente escolhidos, de Wichita e Omaha pedindo a eles que participassem de um estudo de contato social na sociedade americana. A carta continha um breve resumo da proposta do estudo, uma fotografia, nome, endereço e outras informações sobre as duas pessoas escolhidas como alvos, junto com as seguintes instruções:

COMO PARTICIPAR DO ESTUDO

1. adicione o seu nome o da próxima pessoa que receberá esta carta⁹ no final da lista.
2. Pegue um cartão postal, existente em grande quantidade nessa correspondência, e envie de volta a universidade de Harvard. Não é necessário selo. O cartão postal é muito importante. Pois ele nos permitirá acompanhar o progresso desta correspondência ao longo do seu trajeto até a pessoa alvo.
3. Se você conhecer a pessoa alvo, envie esta carta diretamente para ela. Faça isso somente se você conhece a pessoa alvo pelo primeiro nome¹⁰.
4. Se você não conhecer a pessoa alvo, não envie a carta para ela. Ao invés disso, envie essa carta (com cartões postais, fotos e informações sobre os alvos) para alguém que você conheça, e que você acredite que ele conheça a pessoa alvo. Você pode enviar a carta a um amigo, parente ou conhecido, mas deve ser alguém que você conheça pelo primeiro nome.

⁹Na verdade, um pacote contendo as informações das pessoas alvos, uma lista e cartões postais.

¹⁰Para os americanos, tratar (conhecer) uma pessoa pelo primeiro nome, significa que vocês já foram apresentados e de uma forma ou outra mantém algum tipo de relação, que pode ser de coleguismo, amizade, etc.

Enquanto o experimento acontecia, Milgram tinha um pensamento recorrente: será que alguma das cartas chegaria ao alvo? Se o número de ligações necessárias para fazer com que a carta chegasse ao alvo estivesse em torno de 100, como seu amigo pensava, então o experimento provavelmente poderia falhar, uma vez que poderia haver uma pessoa ao longo do caminho que não cooperasse.

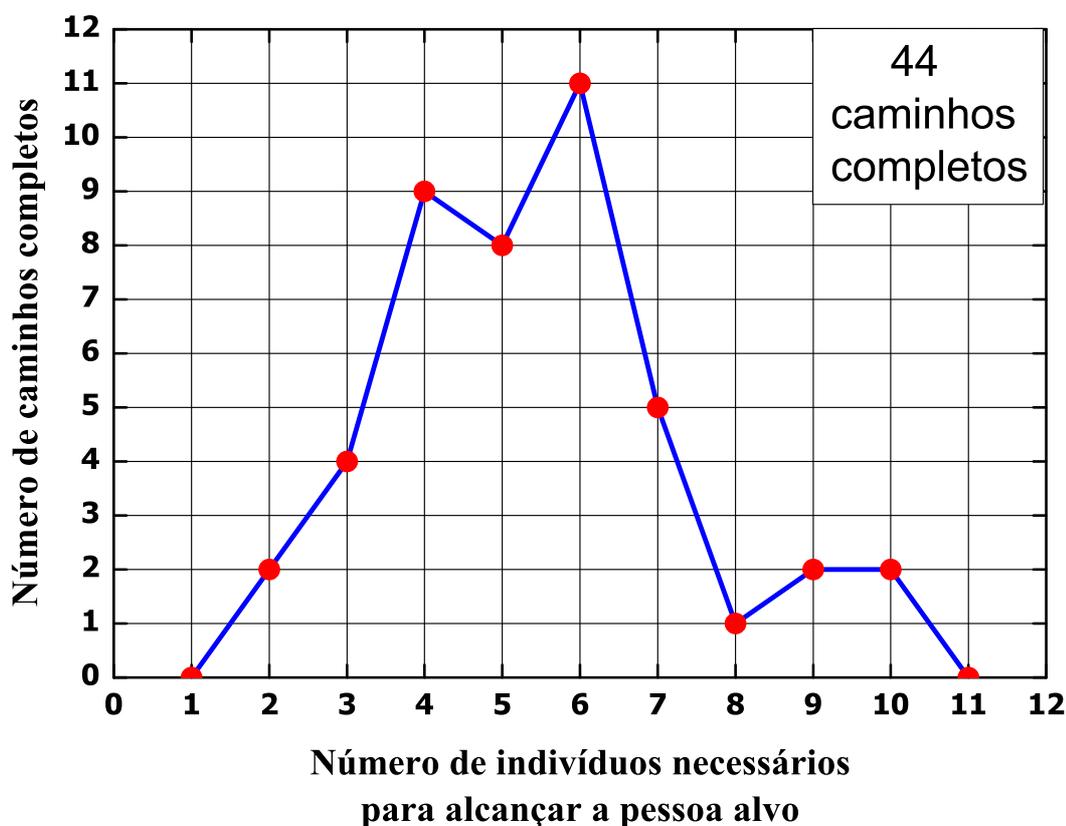


Figura 4.3: Histograma, mostrando que no final do experimento Milgram, as cartas que chegaram ao destino passaram por seis pessoas (em média).

Tamanha foi a surpresa de Milgram, quando poucos dias depois, de iniciado o experimento, a primeira carta chegou ao alvo passando por apenas duas pessoas. Com o passar dos dias mais cartas chegavam ao alvo. De um total de 160 cartas, 44 chegaram ao alvo (veja Figura 4.3), algumas precisaram passar por até 11 pessoas para chegar ao seu destino (o alvo). Estes caminhos completos permitiram a Milgram determinar o número de pessoas necessárias para fazer chegar uma carta ao alvo. Ele encontrou que

o número médio de pessoas, para se fazer uma carta chegar ao alvo, era de aproximadamente 5,5. Um valor muito diferente das 100 pessoas, como havia previsto o amigo de Milgram. Arredondando esse valor de 5,5 para 6, obtemos os famosos “seis graus de separação”. Ao contrário do que muitas pessoas pensam, não foi Stanley Milgram quem cunhou a expressão “seis graus de separação”. Essa expressão teve sua origem numa peça de teatro criado por John Guare em 1991. Depois de fazer grande sucesso na Broadway, essa peça se transformou em filme com o mesmo título “seis graus de separação”. Num trecho do filme, o personagem chamado Ousa (interpretado por Stockard Channing), refletindo sobre a ligação entre as pessoas de todo o mundo diz para sua filha, “Cada um neste planeta está separado apenas por seis outras pessoas. Seis graus de separação. Entre nós e qualquer outra pessoa nesse planeta. O presidente dos Estados Unidos. Um gondoleiro em Veneza.... não necessita nem ao menos ser uma pessoa conhecida. Isso vale para qualquer um. Um nativo de uma floresta tropical. Um habitante da Terra do Fogo. Um esquimó. Eu estou ligado a uma pessoa por um caminho de seis pessoas. É algo muito profundo... como se uma pessoa fosse uma nova porta que se abre para outros mundos.”. Apesar da diferença entre o experimento de Milgram (que diz respeito somente aos Estados Unidos), e a peça de John Guare que dizia respeito a todo o mundo, a idéia dos seis graus de separação (com um apelo científico), nascia naquele momento.

A contribuição de Stanley Milgram não foi apenas nos chamar a atenção para o quanto estamos conectados, mas também demonstrar que ninguém está a muito mais que alguns “apertos de mão” de outra pessoa. Isto é, o mundo é pequeno.

4.4 Ressurgimento da Ciência das Redes

Como é habitual do homem, depois de obtermos as respostas para algumas perguntas, logo surgem novas perguntas. Como, por exemplo, se a distância média de separação entre as duas pessoas quaisquer no mundo é sempre 6. Ou ainda, se tentarmos

formar uma rede com essas pessoas, que estrutura de rede encontraremos? As respostas a essas perguntas começaram a surgir há alguns anos atrás. Hoje em dia, além de sabermos que as redes sociais são mundos pequenos, sabemos um pouco mais. Sabemos como podemos classificá-las, e calcular várias de suas grandezas. Como veremos na subseção 4.13, temos muito mais conhecimento sobre as redes hoje do que quando tudo começou (ou recomeçou) em 1999, com os primeiros artigos de Albert-László Barabási e da sua então aluna de doutorado Réka Albert.

4.5 Redes: definição

Para os físicos (matemáticos), uma rede é um conjunto de itens, que chamamos de vértices (nós), com ligações entre eles, chamadas de conexões (arestas). Essa definição de rede é uma das mais simples, mas existem redes que exigem um tratamento mais cuidadoso. Por exemplo, uma rede pode ter mais de um tipo de vértice, e/ou mais de um tipo de conexão entre os vértices. Esses vértices e conexões podem ser associados a uma variedade de propriedades. No caso de uma rede social, os vértices podem representar homens e mulheres, pessoas de diferentes nacionalidades, localidades, idades, ou quaisquer outras características inerentes ou não, enquanto as conexões, podem representar uma relação de amizade ou hostilidade, uma relação profissional ou qualquer outra relação que possamos imaginar. Tanto os vértices quanto as conexões podem carregar um peso, que por exemplo, diz o “quanto” uma pessoa conhece a outra. Com relação as conexões, elas ainda podem ser direcionadas, ou seja, apontam numa certa direção. Redes compostas de conexões direcionadas são chamadas de redes direcionadas (ou dígrafos, pelos matemáticos). Exemplos de redes direcionadas são: as redes WWW, de citação científica, uma vez que os “links” de páginas na Internet e a citação de artigo seguem uma única direção. Redes direcionadas pode ser cíclicas, quando existir “loops” (laços) ou acíclicas, quando não existir nenhum laço.

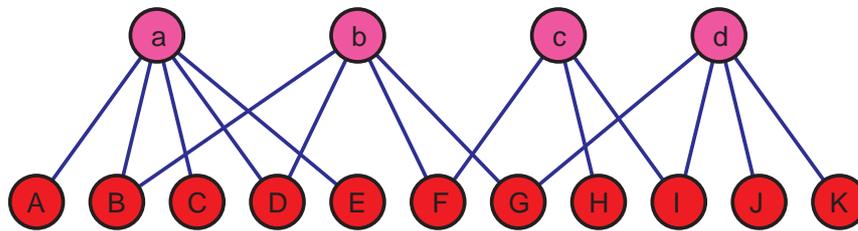


Figura 4.4: Rede bipartida, com 4 vértices do tipo filme e 11 vértices do tipo ator. Vértices de **a** a **d** representam filmes e os vértices de **A** até **K** representam atores, e as linhas ligam os atores aos filmes nos quais eles atuaram.

Existem redes com mais de um tipo de vértice. Nessa tese iremos estudar apenas as redes unipartidas, que tem um único tipo de vértice (veja Figura 4.5) e as bipartidas (veja Figura 4.4), que possuem dois tipos de vértices. Um exemplo de rede bipartida é a rede de atores de Hollywood, que possui dois tipos de vértices (os atores e os filmes) e cada ator possui uma ligação com os filmes nos quais ele atuou. Da rede bipartida podemos construir uma rede unipartida de atores, ligando atores que atuaram juntos num mesmo filme.

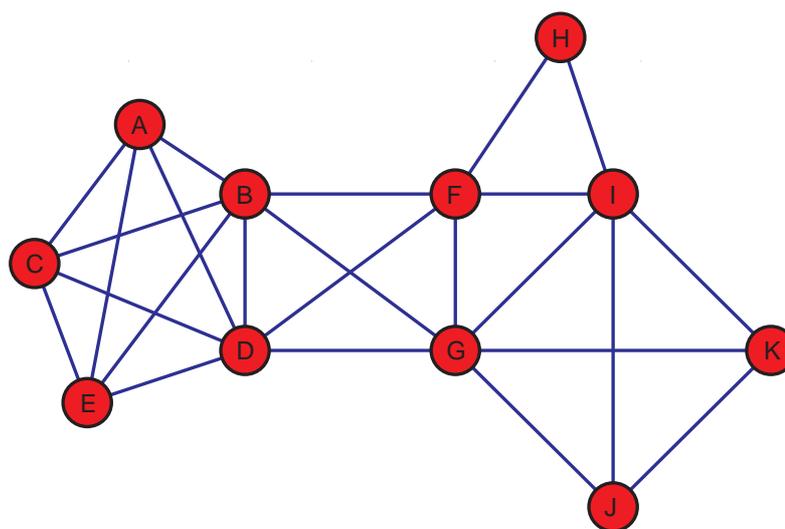


Figura 4.5: Rede unipartida, é a projeção da rede bipartida na qual atores se ligam uns aos outros se eles atuaram num mesmo filme.

O estudo das redes ainda podem levar em conta a evolução temporal, com vértices e ligações aparecendo ou desaparecendo, ou ainda com os mudança dos valores atribuídos aos vértices e ligações. Além disso, existem muitos outros níveis de sofisticação que podem ser adicionados às redes.

Devido ao fato de pesquisadores de várias áreas estarem estudando as redes complexas, muitos termos são usados de maneira equivocada. Para evitar confusão, apresentamos na Tabela 4.1 um pequeno dicionário de termos utilizados nessa tese.

4.6 Tipos de Redes

*P*esquisadores das mais diversas áreas, têm devotado especial atenção ao estudo de redes dos mais diferentes tipos e origem. Muitas vezes, a obtenção de dados dos sistemas reais, é o ponto de partida para o estudo das rede. Nessa tese, iremos seguir a classificação sugerida por Mark E. J. Newman [19] e vamos dividir as redes em quatro categorias: redes sociais, de informação, tecnológicas e biológicas. Nossa intenção é fazer uma breve apresentação das características mais importantes dos diferentes tipos de redes.

Redes Sociais

*U*ma rede social é o conjunto de pessoas ou grupos de pessoas que possuem algum tipo de contato ou interação entre elas. As pessoas fazem o papel de vértice e as interações entre elas são as ligações. As redes sociais seguramente estão entre as mais estudadas [68–71]. A natureza das ligações pode ir desde um contato sexual até um relacionamento profissional entre executivos de empresas. Dois exemplos bastante conhecidos de redes sociais são: rede de atores de Hollywood e a rede de jogadores de futebol brasileiro [33], que veremos em detalhe no Capítulo 6.

Vértice: A unidade fundamental de uma rede, também conhecido como sítio pelos físicos.

Ligação: A linha que conecta dois vértices.

Direcionada/não-direcionada: Uma ligação entre vértices é direcionada se segue numa única direção (como uma rua de mão única), e não-direcionada se ela segue nas duas direções. Uma rede é direcionada se todas ligações são direcionadas.

Conectividade: O número de primeiros vizinhos (vértices adjacentes) do vértice em questão.

Grau: O número de ligações conectados a um vértice. Vale lembrar que o grau não é necessariamente igual ao número de primeiros vizinhos de vértice, uma vez que pode haver mais de uma ligação entre dois vértices. Uma rede direcionada possui os graus de entrada e de saída para cada vértice, que são os números de ligações que entram e saem respectivamente.

Componente: A componente da qual um vértice faz parte é o conjunto de vértices (conectados) que podem ser alcançados seguindo os caminhos das ligações da rede. Vale lembrar que uma rede direcionada, possui componentes de vértices com ligações entrando e de vértices com ligações saindo.

Caminho Geodésico: Mais conhecido como “*shortest path length*” é o caminho geodésico mais curto entre dois vértices quaisquer da rede. É importante dizer que frequentemente existe mais de um caminho geodésico entre dois vértices.

Diâmetro: O diâmetro de uma rede é o caminho geodésico mais longo (dado em termos do número de ligações) entre dois vértices quaisquer da rede.

Tabela 4.1: Definição de alguns dos termos mais utilizados no estudo das redes.

Redes de Informação

*R*edes de informação, também conhecidas como redes de conhecimento são obtidas a partir de bases de conhecimento formal, como as citações de artigos científicos, a World Wide Web, os registros de patentes, a estrutura das linguagens, etc. As citações de artigos científicos são modeladas como um grafo direcionado, praticamente acíclico, dado que as publicações mais recentes citam aquelas mais antigas e nunca o inverso. No outro extremo, a Web forma um grafo direcionado com forte presença de ciclos, uma vez que a evolução temporal permite a um site adicionar ligações a outros de construção posterior.

Redes Tecnológicas

*A*s redes tecnológicas, são redes construídas pelo homem e desenvolvidas para distribuir algum bem de consumo ou algum recurso [72], tal como eletricidade e informação transmitida via Internet [73]. Outros exemplos de redes tecnológicas são: redes de distribuição de eletricidade, água e telefonia.

Redes Biológicas

*S*istemas biológicos podem ser representados como redes. Talvez o exemplo clássico das redes biológicas [74] seja a rede de proteínas de levedura (*yeast*), onde os vértices representam as proteínas e as ligações (dirigidas) representam a interação física entre duas proteínas [75]. Outros exemplos de redes biológicas são os sistemas nervoso e vascular e as cadeias alimentares [76–81], que tem um interesse significativo no estudo de ecossistemas.

4.7 Propriedades das Redes

Um dos modelos mais simples de redes é o modelo de Erdős-Rényi. Neste modelo, ligações são atribuídas aleatoriamente a um número fixo n de vértices, afim de criar uma rede na qual cada uma das $\frac{1}{2}n(n-1)$ possíveis ligações estejam, independentemente, presente com uma probabilidade p , de tal modo que o número de ligações que conecta cada vértice — a conectividade — segue uma distribuição binomial ou uma distribuição de Poisson, no limite de n tendendo para o infinito. Os grafos aleatórios foram estudados inicialmente pelos matemáticos e muitos resultados, alguns aproximados e outros exatos, foram obtidos. Contudo, o que têm atraído a atenção dos pesquisadores nos últimos anos, é o fato de nem todas as redes reais poderem ser representadas como grafos aleatórios.

4.7.1 Distribuição de Conectividades

A conectividade, ou melhor a distribuição de conectividades tem recebido muita atenção nos últimos anos. Essa popularidade se deve aos trabalho de Barabási e seu colaboradores e dos irmãos Faloutsos, mostrando que em algumas redes *reais* a distribuição de conectividades — que pode ser definido como a fração de vértices da rede que tem conectividade k , ou ainda, como a probabilidade de que um vértice escolhido ao acaso tenha conectividade k — decai como uma lei de potência,

$$P(k) = ak^{-\gamma}, \quad (4.1)$$

onde a e γ são constantes. Esse tipo de distribuição está muito longe do que é esperado para uma rede puramente aleatória. A lei de potência tem um papel especial na física estatística, onde ela tem um papel muito importante nos estudos de fenômenos críticos,

e essa é provavelmente uma das razões de tamanha euforia que cerca uma distribuição de conectividades que decai como uma lei de potência. Modelos tradicionais da física estatística usualmente exibem uma transição de fase entre estados ordenados e aleatórios ou vice-versa. Na transição, esse sistemas podem apresentar as propriedades de “auto-similaridade”, “fractalidade” ou “de escala livre”¹¹.

4.7.2 Caminho Geodésico e Diâmetro

Num grafo, conexões que ligam os vértices formam um caminho. Dado dois vértices, pode haver caminhos entre eles, mesmo que não estejam diretamente conectados. O comprimento do menor caminho (do inglês: “*shortest path length*”), ou caminho geodésico, é o caminho mais curto entre dois vértices quaisquer da rede. O diâmetro é o caminho geodésico mais longo entre dois vértices (dado em termos do número de ligações) quaisquer. No caso de um grafo desconectado, *i. e.*, composto por mais de uma componente, o diâmetro é infinito, uma vez que pode não ser possível estabelecer um caminho entre as diversas componentes. Contudo, sempre é possível redefinir o diâmetro do grafo como sendo o diâmetro máximo de seus agrupamentos.

A distância geodésica ℓ média entre pares de vértices de uma rede não-direcionada, pode ser calculada por

$$\ell = \frac{1}{\frac{1}{2}n(n-1)} \sum_{i \geq j} d_{ij} .$$

onde d_{ij} é a distância geodésica entre os vértices i e j .

¹¹O termo “de escala livre” se refere a qualquer forma funcional $f(x)$ que permaneça inalterada na presença de um fator multiplicativo associado a variável independente x . E como apenas as leis de potência respeitam essa condição, $f(ax) = b(a)f(x)$, podemos fazer uso dos termos “lei de potência” e “de escala livre” como sendo sinônimos. Muitas vezes uma distribuição com um comportamento do tipo lei de potência é confundida com as leis de Pareto ou de Zipf (para maiores detalhes veja o Apêndice A)

No caso de uma rede regular d -dimensional (veja Figura 4.6(a)) com N vértices, temos que a distância geodésica é dada por $\ell \sim N^{1/d}$. Enquanto o valor da distância geodésica rede totalmente conexa é sempre 1. Podemos estimar facilmente o valor ℓ de uma rede aleatória 4.6(b), para isso partimos do pressuposto que o número médio de primeiros vizinhos do vértice i é dado por z , então temos que z^ℓ vértices da rede estão a uma distância ℓ do vértice i . Disso temos que, $N \sim z^\ell$ e então $\ell \sim \ln N / \ln z$, *i. e.*, o caminho geodésico médio é pequeno mesmo para grandes redes. Como veremos a seguir, essa característica é conhecida como efeito mundo pequeno.

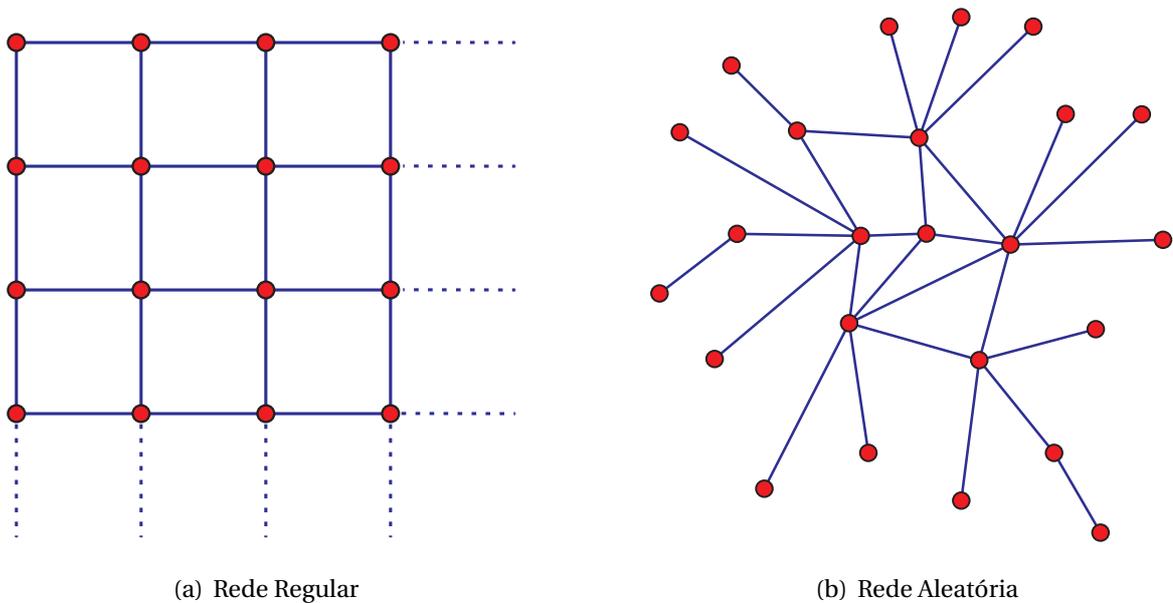


Figura 4.6: Exemplo de dois tipos de redes: rede regular bidimensional 4.6(a) e rede aleatória 4.6(b).

4.7.3 Efeito de Mundo Pequeno

O efeito mundo pequeno, uma propriedade presente em muitos sistemas, significa que apesar de possuírem um grande número de vértices, a distância média entre dois vértices é muito pequena. A manifestação mais popular do efeito “mundo pequeno”

é o conceito de “seis graus de separação”, (re)descoberto pelo psicólogo social Stanley Milgram em 1967, que mostrou a existência de um caminho curto de relacionamento entre duas pessoas quaisquer dos Estados Unidos.

Algumas redes tem distância entre vértices que crescem ainda mais lentamente que $\ln N$. Bollobás e Riordan [82, 83] mostraram que certas redes, com distribuição de conectividades decaindo como uma lei de potência, tem valores de ℓ que não crescem mais rápido do que $\ln N / \ln(\ln N)$, esse efeito foi chamado de “*ultra small-world*”. Cohen e Havlin [84] argumentaram que a variação pode ser ainda mais lenta.

O efeito mundo pequeno existe em vários outros sistemas, como por exemplo na rede de jogadores de futebol brasileiros, onde a distância média entre dois jogadores quaisquer é aproximadamente 3, assim como a rede de reações químicas nas células também tem uma distância de separação de 3.

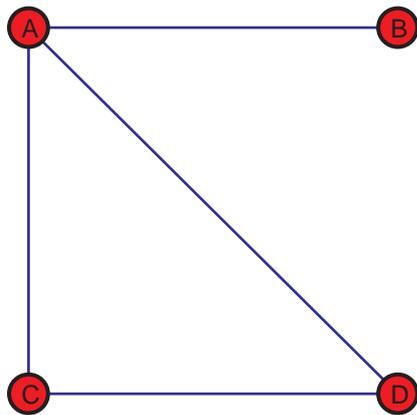
O fato de qualquer um de nós estar a uma curta distância de qualquer outra pessoa no mundo já não é um mistério ou surpresa, afinal de contas vivemos num mundo globalizado onde nossa sociedade é uma gigantesca rede de relacionamentos. E assim como outras redes, a rede de relacionamento também tem dentro de si o efeito mundo pequeno. Sistemas que possuem o efeito mundo pequeno diferem de sistemas euclidianos, onde estamos acostumados a medir distâncias em alguma unidade de comprimento. Nossa habilidade para “alcançar” uma pessoa tem cada vez menos relação com a distância física. “Navegar” nesse novo mundo, não-euclidiano, pode muitas vezes nos confundir, e assim precisamos nos lembrar que existe uma nova “geometria” regendo o nosso mundo e que precisamos compreendê-la. Fazendo isso, teremos a oportunidade de estudar e tentar entender os fenômenos que acontecem a nossa volta.

4.7.4 Matriz Adjacência e o Espectro de Autovalores

Qualquer rede G com N vértices pode ser representada por sua matriz adjacência $A(G)$ (veja Figura 4.7(b)) com $N \times N$ elementos A_{ij} , tal que

$$A_{ij} = A_{ji}, \quad (4.2)$$

e $A_{ij} = 1$ se os vértices i e j estão conectados e 0 (zero) caso contrário.



(a) Rede com 4 vértices e 4 ligações.

| | A | B | C | D |
|---|---|---|---|---|
| A | 0 | 1 | 1 | 1 |
| B | 1 | 0 | 0 | 0 |
| C | 1 | 0 | 0 | 1 |
| D | 1 | 0 | 1 | 0 |

(b) Matriz Adjacência da rede da Figura 4.7(a).

Figura 4.7: Matriz adjacência com sua respectiva rede.

O espectro da rede G é o conjunto de autovalores da sua matriz adjacência $A(G)$. Uma rede com N vértice tem N autovalores λ_j , que são definidos pela densidade espectral

$$\rho(\lambda) = \frac{1}{N} \sum_{j=1}^N \delta(\lambda - \lambda_j), \quad (4.3)$$

O interesse no espectro está no fato da densidade espectral estar relacionado com características topológicas das redes, uma vez que o k -ésimo momento podem ser escritos como

$$M_k = \frac{1}{N} \sum_{j=1}^N (\lambda_j)^k = \frac{1}{N} \text{Tr}(A^k) = \frac{1}{N} \sum_{i_1, i_2, \dots, i_k} A_{i_1 i_2} A_{i_2 i_3} \dots A_{i_k i_1}. \quad (4.4)$$

O número de caminhos (ciclos) que retornam ao vértice inicial depois de k passos é $N M_k$.

Esses caminhos podem conter vértices que já foram visitados.

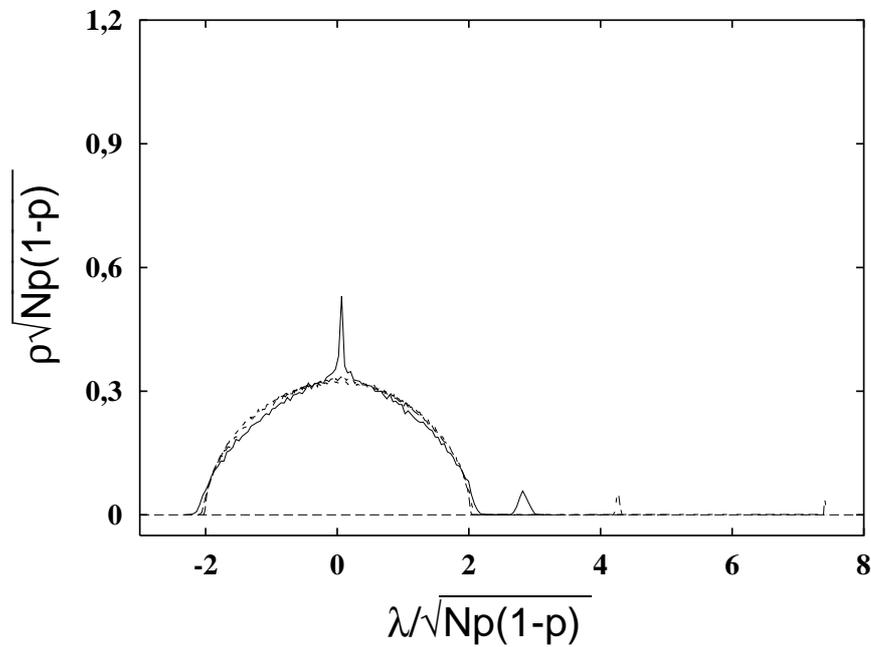


Figura 4.8: Densidade espectral de três ER, com $p = 0,05$ e tamanhos $N = 100$ (linha contínua), $N = 300$ (linha tracejada) e $N = 1000$ (linha pontilhada). [85]

Na Figura 4.8, vemos a densidade espectral de uma rede aleatória, que segue a lei de Wigner, também conhecida como lei do semicírculo, dada por

$$\rho(\lambda) = \begin{cases} \frac{\sqrt{4Np(1-p) - \lambda^2}}{2\pi Np(1-p)} & \text{se } |\lambda| < 2\sqrt{Np(1-p)} \\ 0 & \text{caso contrário} \end{cases}. \quad (4.5)$$

4.7.5 Coeficiente de Agrupamento ou Transitividade

O coeficiente de agrupamento de um vértice [31, 61, 62, 86] ou transitividade, quantifica a densidade de conexões próximo a um dado vértice. Essa tendência ao agrupamento é quantificada pelo coeficiente de agrupamento.

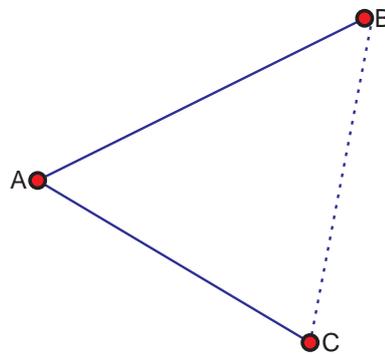


Figura 4.9: Rede social composta de três pessoas, onde **A** é amigo de **B** e de **C**. Nessas redes, existe uma grande chance de **B** ser amigo de **C**.

Podemos definir qualitativamente o coeficiente de agrupamento da seguinte forma: se o vértice **A** está diretamente ligado nos vértices **B** e **C**, então existe uma probabilidade de que **B** se conecte ao vértice **C**, veja Figura 4.9. Em termos de topologia de redes, transitividade significa uma presença elevada de triângulos.

Quantitativamente o coeficiente de agrupamento do vértice i é definido da seguinte forma: seja i um vértice na rede com k_i ligações que conectem ele a k_i outros vértices. O valor do coeficiente de agrupamento do vértice i é dado pela razão entre o número E_i de ligações que existem entre o vértice i e seus k_i primeiros vizinhos e o número total de ligações possíveis entre o vértice i e seus k_i primeiros vizinhos, $k_i(k_i - 1)/2$,

$$C_i = \frac{2E_i}{k_i(k_i - 1)}, \quad (4.6)$$

que assume valores no intervalo $0 \leq C_i \leq 1$. O coeficiente de agrupamento de toda a rede, coeficiente de agrupamento médio, é obtido fazendo a média do C_i de cada vértice

$$C = \frac{1}{N} \sum_i C_i = \frac{1}{N} \sum_i \frac{2E_i}{k_i(k_i - 1)}. \quad (4.7)$$

Para exemplificar o conceito de coeficiente de agrupamento da Equação 4.6, vemos na Figura 4.10 uma rede com 5 vértices. Nessa figura vemos três redes e os respectivos coeficientes de agrupamentos do vértice azul (central).

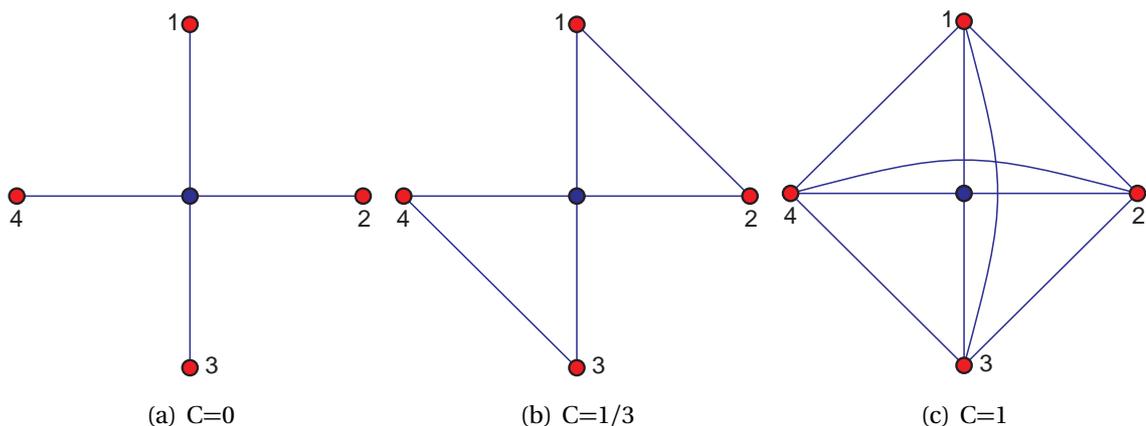


Figura 4.10: Ilustração da definição do coeficiente de agrupamento (veja Equação 4.6). O coeficiente de agrupamento do vértice azul, localizado no centro da rede, para as três redes. Na primeira, o valor do coeficiente de agrupamento é **zero**, pois nenhum vizinho do vértice azul tem ligação entre si. Na segunda, quando existem duas ligações entre vizinhos do vértice azul, o coeficiente de agrupamento vale **1/3** e na terceira quando todos os vizinhos do vértice azul se conectam, o coeficiente de agrupamento é **1**, o valor máximo.

As rede aleatórias tem um valor de coeficiente de agrupamento muito baixo. A razão está no fato das ligações serem distribuídas aleatoriamente, assim a probabilidade de que os primeiros vizinhos de um vértice estejam conectados é p , a mesma de que quaisquer outros dois vértices na rede estejam conectados, conseqüentemente o coeficiente de agrupamento de um grafo aleatório é dado por

$$C = p = \frac{\langle k \rangle}{N} \quad (4.8)$$

onde $\langle k \rangle$ é a conectividade média e N é o número de vértices da rede aleatória.

Como veremos na Seção 4.12, o cenário do estudo das redes começou a mudar quando Duncan Watts e Steven Strogatz em 1998 sugeriram que em muitas redes reais o coeficiente de agrupamento é tipicamente muito maior do que num grafo aleatório com o mesmo número de vértices e ligações. Esta foi uma das primeiras observações feitas indicando que as redes reais tinham propriedades que iam além dos grafos aleatórios.

4.7.6 Correlação e Conectividade Média dos Primeiros Vizinhos

Como já vimos, a distribuição de conectividades é uma grandeza importante para caracterizar redes complexas. Uma outra grandeza muito útil na caracterização das redes é a correlação de conectividade. Uma rede é dita correlacionada se a conectividade de seus vértices dependem da conectividade dos seus vizinhos.

Para verificar se uma rede é correlacionada, ou não, fazemos uso da probabilidade condicional $P(k'|k)$ — probabilidade de que uma ligação de um vértice de conectividade k esteja ligado a um outro vértice de conectividade k' — tal que $P(k'|k)$ satisfaz a normalização

$$\sum_{k'} P(k'|k) = 1, \quad (4.9)$$

e a condição de balanço detalhado [87], dada por

$$kP(k'|k)P(k) = k'P(k|k')P(k'). \quad (4.10)$$

Para redes não correlacionadas, onde $P(k'|k)$ não depende de k , a probabilidade condicional pode ser reescrita como

$$P(k'|k) = \frac{k'P(k')}{\langle k \rangle}. \quad (4.11)$$

Embora a correlação de conectividade seja formalmente caracterizada por $P(k'|k)$, podemos estimar correlação de conectividade do vértice i de uma forma bem simples

$$k_{nn,i} = \frac{1}{k_i} \sum_{j \in N_i} k_j = \frac{1}{k_i} \sum_{j=1}^N A_{i,j} k_j, \quad (4.12)$$

onde k_i é o número de vizinhos do vértice i e N_i é o conjunto de primeiros vizinhos do vértice i e lembrando da Seção 4.7.4, $A_{i,j}$ é a matriz adjacência da rede. Na Figura 4.11, vemos uma rede com 19 vértices que nos auxiliará a compreender como é feito o cálculo da correlação de conectividade do vértice i . Fazendo uso da Equação 4.12, temos que $k_{nn,i} = 1/4 \times (3 + 4 + 4 + 7) = 18/4$.

Em 2001 Pastor-Satorras *et al.* [88], propuseram o cálculo da conectividade média dos primeiros vizinhos de um vértice de conectividade k , $\bar{k}_{nn}(k)$, em função da probabilidade condicional da seguinte forma,

$$\bar{k}_{nn}(k) = \sum_{k'} k' P(k'|k). \quad (4.13)$$

Se não existe correlação entre os vértices, fazendo uso da Equação 4.11 podemos reescrever a Equação 4.13 como

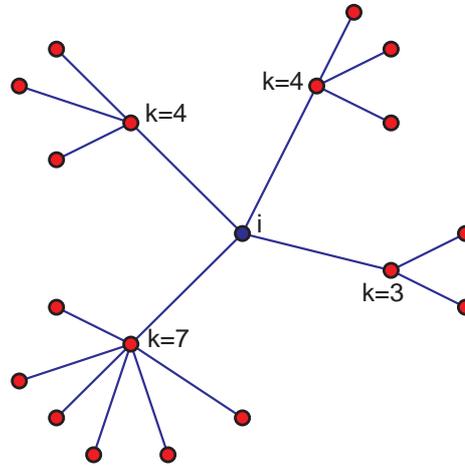


Figura 4.11: Rede com 19 vértices, utilizada para exemplificar o cálculo da correlação de conectividade $k_{nn,i}$.

$$\bar{k}_{nn}(k) = \sum_{k'} k' \left[\frac{k' P(k')}{\langle k \rangle} \right] = \frac{\langle k^2 \rangle}{\langle k \rangle}, \quad (4.14)$$

ou seja, $\bar{k}_{nn}(k)$ é independente de k .

Redes correlacionadas são classificadas como assortativa (que traduzimos do inglês, *assortative*) se $\bar{k}_{nn}(k)$ é uma função crescente de k , são classificadas como desassortativa (que traduzimos do inglês, *disassortative*) se $\bar{k}_{nn}(k)$ é uma função decrescente de k e são classificadas como não assortativa (que traduzimos do inglês, *no assortative*) se $\bar{k}_{nn}(k)$ for uma função constante. Em outras palavras, em redes assortativas os vértices tendem a se conectar com outros de conectividade semelhantes, nas redes desassortativas vértices de baixa conectividade tendem a se conectar com vértices de alta conectividade e em redes não assortativas todos os vértices têm conectividade média semelhante.

Em várias redes complexas a grandeza conectividade média dos primeiros vizinhos, $\bar{k}_{nn}(k)$, decai como uma lei de potência. Essa grandeza tem sido quantificada pelo expoente ν . Nas Figuras 4.12 e 4.13, vemos as correlações de conectividades $\bar{k}_{nn}(k)$ versus k de algumas redes de sistemas reais.

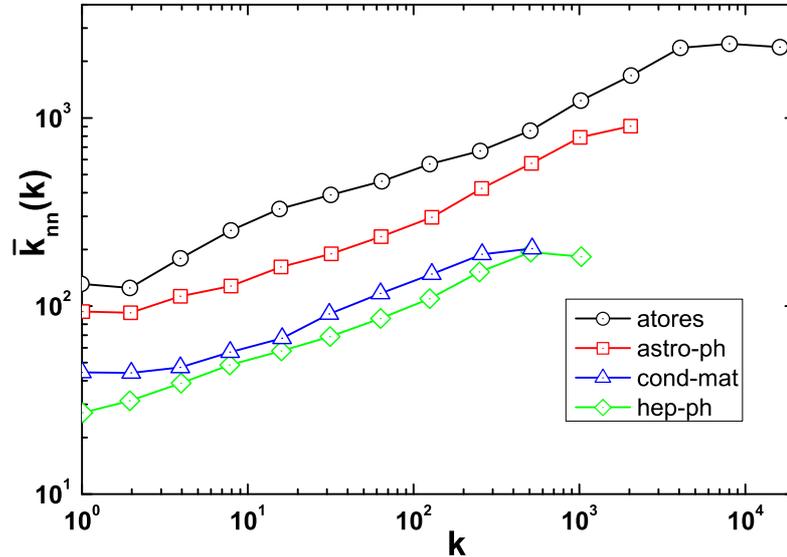


Figura 4.12: Conectividade média dos primeiros vizinhos, $\bar{k}_{nn}(k)$, das redes de atores, de citações das áreas de astrofísica, física da matéria condensada e física de altas energias. Todas as quatro redes, seguem uma lei de potência com expoente $\nu \approx 0,3$.

Na Figura 4.12 são apresentadas a conectividade média dos primeiros vizinhos das redes de atores, de citações das áreas de astro-física, física da matéria condensada e física de altas energias. Todas as quatro redes são correlacionadas, ou seja, são assortativas. Essa característica pode ser facilmente verificada pela função crescente $\bar{k}_{nn}(k)$ versus k das quatro redes.

Na Figura 4.13 vemos conectividade média dos primeiros vizinhos das redes: de roteadores da Internet (IR) e dos sistemas autônomos (AS) e AS+, onde AS+ são sistemas autônomos “enriquecidos” com relação ao número de ligações. Apesar da diferença de representação, as redes IR e AS descrevem a Internet em diferentes escalas, onde cada vértice da rede AS é um conjunto de roteadores da Internet.

A correlação de conectividade também pode ser estimada pelo coeficiente de correlação r de Pearson, que é calculado da seguinte forma

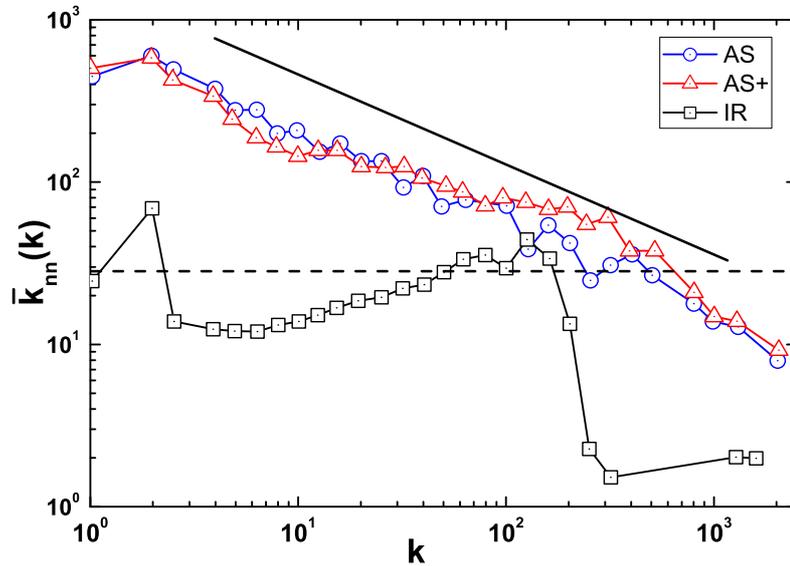


Figura 4.13: Conectividade média dos primeiros vizinhos das redes IR, AS e AS+. A linha sólida é dada pelo decaimento de lei de potência $\bar{k}_{nn}(k) \sim k^{-0,55}$, indicando que as redes AS e AS+ são correlacionadas e desassortativas. A linha horizontal pontilhada marca o valor da ausência de correlação $\bar{k}_{nn}(k) = \langle k^2 \rangle / \langle k \rangle = 26,9$, ou seja a rede IR é não assortativa.

$$r = \frac{\sum_{k=1}^N (A_{ik} - \bar{A}_i)(A_{jk} - \bar{A}_j)}{\sqrt{\sum_{k=1}^N (A_{ik} - \bar{A}_i)^2} \sqrt{\sum_{k=1}^N (A_{jk} - \bar{A}_j)^2}}, \quad (4.15)$$

onde A_{ik} é a matriz adjacência da rede em questão e \bar{A}_i é o valor médio da linha i da matriz adjacência. O resultado da Equação 4.15 é um simples número real entre -1 e 1 . Se $0 < r \leq 1$, então a rede é dita assortativa. No caso de $-1 \leq r < 0$, a rede é dita desassortativa e se r for 0 (zero) a rede é classificada como não assortativa.

4.8 Erdős e Rényi e os Grafos Aleatórios

Entre os matemáticos, uma rede é mais conhecida como sendo um grafo.

Um grafo é o conjunto de pares $G = \{P, E\}$, onde P é o conjunto de N vértices (P_1, P_2, \dots, P_N) .

e E é um conjunto de ligações que conectam dois elementos de P .

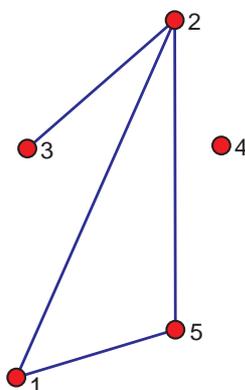


Figura 4.14: Grafo com $N = 5$ vértices e 4 ligações. O conjunto de vértices é $P = \{1, 2, 3, 4, 5\}$, e o conjunto de ligações é $E = \{\{1, 2\}, \{1, 5\}, \{2, 3\}, \{2, 5\}\}$.

Como já vimos na subseção 4.2, a teoria de grafos teve origem no início do século XVIII com os trabalhos de Leonard Euler. Contudo, foi apenas no século XX que a teoria de grafos passou a interessar mais os matemáticos e físicos estatísticos. A primeira tentativa de construir um modelo para redes grandes (e aparentemente) aleatórias foi feita por R. Solomonoff e Anatol Rapoport [89]. Alguns anos mais tarde, e de maneira totalmente independente, Erdős e Rényi [58] redescobriram e estudaram exhaustivamente, algo que chamaram de “grafos aleatórios” (*random graph*), que é como conhecemos até hoje.

No artigo primeiro sobre grafos, intitulado *On Random Graphs* [58], Erdős e Rényi definem um grafo aleatório da seguinte forma: dado um número N de vértices, conecte (não conecte) cada par com uma probabilidade p ($1 - p$). Isto define o modelo Erdős-Rényi conhecido como $G_{N,p}$. Esse modelo é o *ensemble* de todos os possíveis grafos que podem ter m ligações, com probabilidade $p^m(1 - p)^{M-m}$, onde $M = \frac{1}{2}N(N - 1)$ é o número máximo de ligações.

Podemos expressar as propriedades de $G_{N,p}$ não em termos de p mas em termos da conectividade média z de um vértice. Tendo que o número médio de ligações no grafo é $\frac{1}{2}N(N - 1)p$, e somando as conectividades de todos os vértices, temos que

$$\sum_{k=1}^{\infty} k N(k) = 2 \times \# \text{ de ligações} = 2 \times \frac{N(N-1)p}{2} = N(N-1)p. \quad (4.16)$$

onde N_k é o número de vértice de conectividade k e $\sum_k N_k = N$. Fazendo uso da Equação 4.16 podemos escrever a conectividade média z de um vértice como sendo

$$z = \frac{1}{N} \sum_{k=1}^{\infty} k N(k) = \frac{N(N-1)p}{N} = (N-1)p \simeq Np. \quad (4.17)$$

A probabilidade $P(k)$ de que um vértice tenha exatamente k , na rede Erdős- Rényi, é dada pela distribuição binomial:

$$P(k) = \binom{N}{k} p^k (1-p)^{N-k}, \quad (4.18)$$

reescrevendo a distribuição $P(k)$ no limite de $p \rightarrow 0$ e $Np = z$, onde z é uma constante, temos

$$P(k) = \frac{z^k e^{-z}}{k!}, \quad (4.19)$$

A estrutura (a forma como os vértices se ligam) de um grafo varia muito com o valor de p , e certamente uma das mais importantes propriedades dos grafos aleatórios é a transição de fase. Numa fase de baixa densidade, quando p assume valores pequenos, fazendo com que existam poucas ligações entre os vértices, por isso, a rede é composta por várias componentes pequenas, com distribuição de tamanhos de componentes decaindo exponencialmente e um tamanho médio de componente finito. Quando p assume valores grandes, a maior parte dos vértices se juntam criando uma componente

gigante¹². Com exceção da componente gigante, o restante dos vértices que pertencem a componentes menores possuem distribuição de tamanhos de componentes decaindo exponencialmente e um tamanho médio de componente finito. Erdős e Rényi demonstraram, em uma série de artigos [58–60], que é possível resolver exatamente algumas das propriedades dos grafos aleatórios quando assumimos N (número de vértices do grafo) muito grande.

Podemos, por exemplo, demonstrar que o grafo aleatório possui uma transição de fase no tamanho de componentes, onde uma componente é um subconjunto de vértices do grafo conectados apenas entre si. Para um valor pequeno de z , ou seja, quando existem poucas ligações no grafo, encontramos muitos vértices desconectados, as componentes existentes são pequenas, e mantém um tamanho médio constante a medida que o grafo cresce. Entretanto, existe um valor crítico de z acima do qual a maior componente do grafo contém uma fração finita S do número total de vértices. Essa componente maior do grafo é chamada de componente gigante. O valor de z , no qual acontece a transição de fase e surge a componente gigante é precisamente $z = 1$. O parâmetro de ordem para a transição de fase do tamanho de componentes num grafo é a fração de vértices S do grafo que ocupa a componente gigante.

Para calcular o tamanho da componente gigante, podemos fazer uso do seguinte argumento heurístico: seja u a fração de vértices no grafo que não pertença a componente gigante, que também podemos interpretar como sendo a probabilidade de que um vértice escolhido aleatoriamente na rede não pertença a componente gigante. Por sua vez, esta também é igual a probabilidade de que nenhum dos vizinhos desse vértices pertença a componente gigante, que é dado por u^k se o vértice tem conectividade k . Fazendo a média dessa expressão, usando a distribuição de probabilidade $P(k)$ da Equação 4.18, obtemos a relação auto-consistente

¹²Também conhecida como componente simples ou isolada.

$$u = \sum_{k=0}^{\infty} P(k)u^k = e^{-z} \sum_{k=0}^{\infty} \frac{(zu)^k}{k!} = e^{z(u-1)}. \quad (4.20)$$

A fração S do grafo ocupado pela componente gigante é $S = 1 - u$, e fazendo uso da Equação 4.20, a fração S pode ser reescrita como

$$S = 1 - e^{-zS}. \quad (4.21)$$

A Equação 4.21 é transcendental, mas é possível ver que para $z < 1$ a única solução não negativa é $S = 0$, enquanto para $z > 1$ existe uma solução diferente de zero, que é o tamanho da componente gigante. A transição de fase acontece quando $z = 1$. Como podemos ver na Figura 4.15, esse também é o ponto de divergência do tamanho médio das componentes.

Usando funções geratrizes (veja Seção 4.10), podemos escrever o tamanho médio $\langle s \rangle$ da componente ao qual pertence um vértice aleatoriamente escolhido (excluindo a componente gigante), como sendo

$$\langle s \rangle = \frac{1}{1 - z + zS}. \quad (4.22)$$

Para construirmos a curva de tamanho médio $\langle s \rangle$ das componentes, como na Figura 4.15, devemos fazer o seguinte: substituir na Equação 4.21 valores de z , da qual obteremos como saída valores de S , e então substituir esses pares de z e S na Equação 4.21 de onde obteremos como resposta valores de tamanhos médios $\langle s \rangle$ das componentes.

Muito do nosso conhecimento sobre as redes veio dos estudos das redes aleatórias. Sem dúvida, a presença da transição de fase e a existência da componente gigante são idéias que fundamentam e permeiam grande parte destes estudos. Muitas

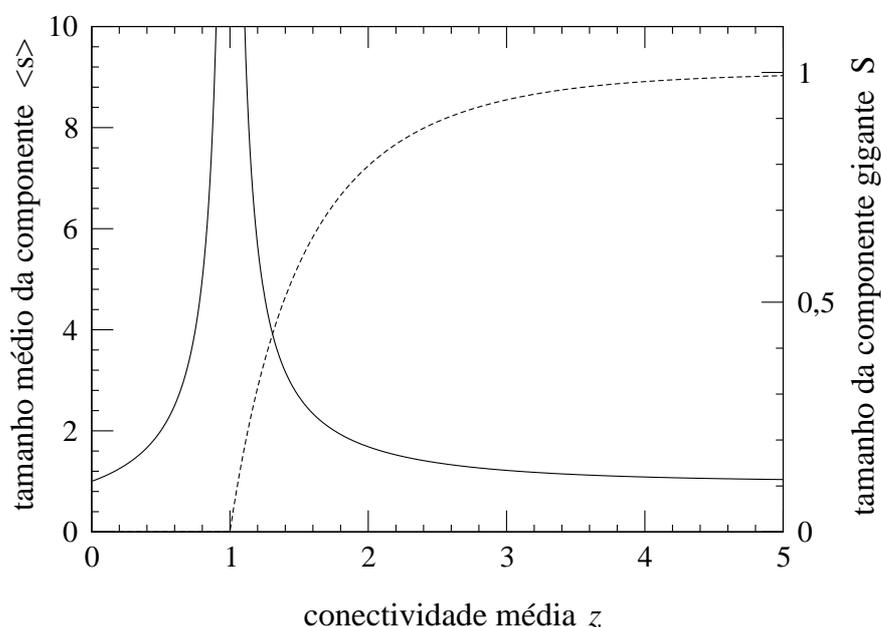


Figura 4.15: Representação da evolução do tamanho médio das componentes (linha contínua) e do tamanho da componente gigante (linha pontilhada) *vs.* a conectividade média, no caso de um grafo aleatório de Erdős-Rényi.

propriedades das redes aleatórias como a distribuição de conectividades, o coeficiente de agrupamento, a conectividade média de primeiros vizinhos entre outras, não reproduzem características de redes do mundo real.

As redes aleatórias, assim como as redes reais, apresentam o efeito mundo pequeno. Uma importante diferença entre essas redes é que, enquanto as redes reais, possuem um valor coeficiente de agrupamento alto, as redes aleatórias possuem um valor de coeficiente de agrupamento baixo. Como veremos na Seção 4.12, devemos a Watts e Strogatz, a criação de uma rede que possua simultaneamente valor alto do coeficiente de agrupamento e o efeito de mundo pequeno.

4.9 Modelo de Configuração

Chama-se modelo de configuração (do inglês, *configuration model*), a rede aleatória com uma dada (e fixa) distribuição de conectividades. Proposto inicial-

mente por Bender e Canfield [90] em 1978, o modelo foi criado para gerar grafos aleatórios que não possuíssem uma distribuição de conectividades poissoniana. A idéia do modelo é se restringir a uma seqüência de conectividades, isto é, a um conjunto específico $\{k_i\}$ de conectividades dos vértices $i = 1, \dots, N$. Esse conjunto é escolhido de tal modo que a fração de vértices que tem conectividade k tenda a uma desejada distribuição $P(k)$ quando o número vértices N é suficientemente grande. No caso de uma simulação numérica, podemos simplesmente sortear a seqüência de conectividades $\{k_i\}$ de uma distribuição $P(k)$ desejada.

De posse da seqüência de conectividades, os passos para gerar grafos aleatórios com uma dada distribuição de conectividades são os seguintes: cada vértice i tem um número de ligações k_i de ligações incompletas (a serem completadas), e então escolhemos aleatoriamente pares dessas ligações incompletas e as unimos formando ligações completas. Quando não existir mais nenhuma ligação incompleta o grafo resultante é uma amostra possível do *ensemble* de grafos, com a seqüência de conectividades desejada.

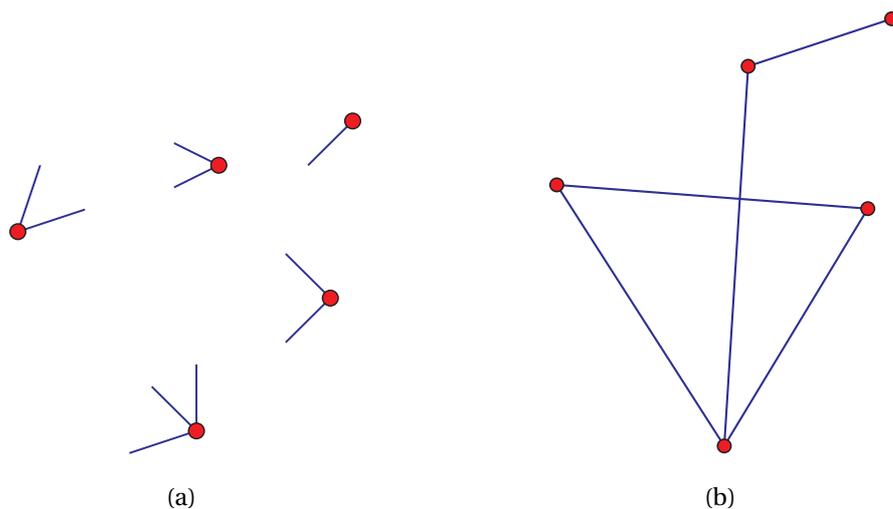


Figura 4.16: Construção de uma rede aleatória com uma dada distribuição de conectividades $P(k)$. (a) Conjunto de $\{k_i\}$ ligações incompletas. (b) Rede formada conectando aleatoriamente pares de $\{k_i\}$ ligações incompletas.

Uma ferramenta muito útil no modelo de configuração são as funções geratrizes. Por essa razão, deste ponto até o fim desta seção faremos uma breve revisão formalismo da função geratriz. Para isso iremos nos valer do valioso, e muito bem escrito, artigo de Mark E. J. Newman intitulado “*Random Graphs as Models of Networks.*”¹³.

Para começar, suponha que estamos interessados no número médio de segundos vizinhos de um vértice, isto é, o número médio de vértices que estão a dois passos (duas ligações) de distância de um vértice. Numa rede que possua agrupamento, muitos dos segundos vizinhos de um vértice também são seus primeiros vizinhos. Contudo, numa rede aleatória, a probabilidade de que um dos segundos vizinhos de um vértice seja também um de seus primeiros vizinhos escala com N^{-1} , e sendo assim, pode ser ignorado no limite de N tendendo ao infinito.

Para calcular corretamente o número de segundos vizinhos numa rede aleatória, precisamos nos lembrar que $P(k)$ é a distribuição de conectividades de vértices de uma rede, mas a conectividade de um vértice que alcançamos seguindo uma ligação aleatoriamente escolhida na rede não é dada por $P(k)$. Visto que existem k ligações chegando num vértice de conectividade k , é k vezes mais provável chegarmos num outro vértice de conectividade igual ao daquele de onde saímos do que chegar num vértice de conectividade 1. Temos então que a distribuição de conectividades do vértice, que encontramos sorteando aleatoriamente uma ligação e seguindo para um lado, é proporcional a $kP(k)$.

É importante frisar essa diferença, pois na maioria dos casos, não estamos interessados na conectividade completa de um vértice, mas no número de ligações que emergem de um vértice exceto aquela por onde chegamos, visto que a ligação por onde chegamos não contribui para o número de segundos vizinhos do vértice escolhido. Chamamos esse número de conectividade excedente (do inglês, *excess degree*), que é a conec-

¹³Esse artigo encontra-se disponível na internet no repositório **arxiv** sob o código cond-mat/0202208. Vale lembrar também que este artigo é um capítulo integrante do livro *Handbook of Graphs and Networks*, publicado por S. Bornholdt e H. G. Schuster.

tividade total de um vértice menos 1 (um). A conectividade excedente tem uma distribuição $Q(k)$ dada por

$$Q(k-1) = \frac{k P(k)}{\sum_j j P(j)} = \frac{k P(k)}{z}, \quad (4.23)$$

onde $k \geq 1$ e $z = \sum_j j P(j)$ é o fator de normalização da distribuição e a conectividade média dos vértices da rede. Equivalentemente temos que a conectividade excedente pode ser escrita como

$$Q(k) = \frac{(k+1)P(k+1)}{\sum_j j P(j)} = \frac{(k+1)P(k+1)}{z}, \quad (4.24)$$

e a conectividade média desse vértice é dada por

$$\sum_{k=0}^{\infty} k Q(k) = \frac{\sum_{k=0}^{\infty} k(k+1)P(k+1)}{\sum_j j P(j)} = \frac{\sum_{k=0}^{\infty} (k-1)k P(k)}{\sum_j j P(j)} = \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle}. \quad (4.25)$$

O resultado da Equação 4.25, é o número médio de segundos vizinhos de um vértice. Ao multiplicarmos esse valor pela conectividade média de um vértice, que é $z = \langle k \rangle$, encontramos que o número médio de segundos vizinhos de um vértice é

$$z_2 = \langle k^2 \rangle - \langle k \rangle. \quad (4.26)$$

É possível estender o cálculo do número de vizinhos vértices ainda mais distantes. Basta lembrar que, o número médio de ligações chegando nos segundos vizinhos, exceto aquele pelo qual chegamos, também é dado pela Equação 4.25, e isso é verdade para qualquer distância m de um vértice. Dessa forma, o número médio de vizinhos a uma distância m é dado por

$$z_m = \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} z_{m-1} = \frac{z_2}{z_1} z_{m-1}, \quad (4.27)$$

onde $z_1 \equiv z = \langle k \rangle$ e z_2 é dado pela Equação 4.26, e iterando essa equação obtemos que

$$z_m = \left[\frac{z_2}{z_1} \right]^{m-1} z_1. \quad (4.28)$$

Da Equação 4.28, vemos que dependendo de z_2 e z_1 , essa expressão divergirá ou convergirá exponencialmente com o crescimento de m . O número total de vizinhos de um vértice em todas distâncias é finito se $z_2 < z_1$, e é infinito se $z_2 > z_1$ (no limite de N infinito). Ou seja, se a Equação 4.28 tiver como resultado um número finito, provavelmente não existe uma componente gigante no grafo, e se o resultado for um número infinito, então há uma grande probabilidade que exista uma componente gigante no grafo.

No caso do grafo **ER**, a Equação 4.26 é equivalente à $\langle k^2 \rangle - 2\langle k \rangle = 0$, e podemos escrevê-la da seguinte forma:

$$\sum_{k=0}^{\infty} k(k-2)p_k = 0. \quad (4.29)$$

Molloy e Reed [92], foram os primeiros a obter essa condição da posição da transição de fase no grafo aleatório com uma dada distribuição de conectividades.

Poderíamos continuar analisando a rede usando o método empregado até o momento, contudo existe a ferramenta matemática conhecida como probabilidade função geratriz (do inglês, *probability generating function*), que é mais elegante e que nos economiza um trabalho algébrico árduo.

4.10 Probabilidade Função Geratriz

Nessa seção, apresentamos a definição e as propriedades de funções geratrizes, bem como algumas aplicações das funções geratrizes na análise de redes aleatórias.

Uma função geratriz tem a forma:

$$G(x) = g_0 + g_1x + g_2x^2 + \dots = \sum_{n \geq 0} g_n x^n. \quad (4.30)$$

onde é dito que $G(x)$ (ou apenas G) é a função geratriz para a seqüência $\langle g_0, g_1, g_2, \dots \rangle$.

Tomando a distribuição de probabilidade $P(k)$ introduzida na seção anterior, temos

$$G_0(x) = \sum_{k=0}^{\infty} P(k)x^k. \quad (4.31)$$

Essa função captura todas as informações das distribuição original $P(k)$, e para recuperar $P(k)$ de $G_0(x)$ basta fazer

$$P(k) = \frac{1}{k!} \left. \frac{d^k G_0}{dx^k} \right|_{x=0}. \quad (4.32)$$

Assim, podemos dizer que a função G_0 “gera” a distribuição de probabilidade $P(k)$.

Usando a notação de função geratriz, podemos a função geratriz da distribuição $Q(k)$ (veja Equação 4.24) de excesso de ligações que saem do vértice que alcançamos seguindo uma ligação na rede

$$G_1(x) = \sum_{k=0}^{\infty} Q(k)x^k = \frac{\sum_{k=0}^{\infty} (k+1)P(k+1)x^k}{\sum_j j p_j} = \frac{\sum_{k=0}^{\infty} kP(k)x^{k-1}}{\sum_j j p_j} = \frac{G'_0(x)}{z}, \quad (4.33)$$

sendo $G'_0(x)$ a derivada primeira de $G_0(x)$ com respeito ao argumento x . Como veremos na Seção 4.11, as funções geratrizes $G_0(x)$ e $G_1(x)$ serão de grande utilidade.

4.10.1 Propriedades da Função Geratriz

Veremos que, ao longo deste capítulo utilizaremos algumas das propriedades das funções geratrizes. A primeira propriedade é

$$G_0(1) = \sum_{k \geq 0} P(k) = 1, \quad (4.34)$$

uma vez que $\sum_{k=0}^{\infty} P(k) = 1$, então $G_0(x)$ é convergente para todo $|x| \leq 1$.

A segunda propriedade é o cálculo da conectividade média, $\langle k \rangle$, que é feito simplesmente diferenciando $G_0(x)$, com $x = 1$

$$G'_0(1) = \sum_k kP(k) = \langle k \rangle. \quad (4.35)$$

Além do primeiro momento, ou conectividade média, qualquer outro momento da distribuição pode ser obtido, basta fazer a derivada adequada. No caso geral temos que,

$$\langle k^n \rangle = \sum_k k^n P(k) = \left[\left(x \frac{d}{dx} \right)^n G_0(x) \right]_{x=1}. \quad (4.36)$$

A terceira, e mais importante, é a propriedade multiplicativa que diz o seguinte: se uma função geratriz “gera” a distribuição de probabilidade de uma propriedade k de um objeto, como por exemplo a conectividade de um vértice, então a soma dessa propriedade sobre n é distribuída de acordo com a n -ésima potência da função geratriz. A soma das conectividades dos n vértices escolhidos aleatoriamente numa rede tem uma distribuição gerada pela função $[G_0(x)]^n$. Podemos ver isso, notando que o coeficiente do termo x^m em $[G_0(x)]^n$ tem um termo da seguinte forma $P(k_1)P(k_2) \dots P(k_n)$ para cada conjunto $\{k_i\}$ de conectividades, tal que $\sum_i k_i = m$, ou seja, a soma das conectividades dos n vértices é m . Para ilustrar essa propriedade, consideramos um caso onde apenas dois vértices são sorteados, então a função geratriz utilizada é a $[G_0(x)]^2$, que pode ser desenvolvida da seguinte forma

$$\begin{aligned} [G_0(x)]^2 &= \left[\sum_k P(k)x^k \right]^2 & (4.37) \\ &= \sum_{k,j} P(j)P(k)x^{j+k} \\ &= P(0)P(0)x^0 + (P(0)P(1) + P(1)P(0))x^1 \\ &\quad + (P(0)P(2) + P(1)P(1) + P(2)P(0))x^2 \\ &\quad + (P(0)P(3) + P(1)P(2) + P(2)P(1) + P(3)P(0))x^3 + \dots \end{aligned}$$

O coeficiente da potência x^m desta equação é exatamente a soma de todos os produtos $P(j)P(k)$ tal que $j + k = m$, que resulta na probabilidade de que a soma das conectividades de dois vértices seja m . Muitos dos resultados que serão apresentados dependem dessa terceira propriedade, e esse é o motivo da função geratriz ser tão útil no estudo das redes aleatórias.

Para entendermos um pouco melhor as propriedades da função geratriz, vamos fazer uso de alguns exemplos específicos. No caso da rede aleatória de Erdős-Rényi, temos que a função geratriz para a distribuição de conectividades é dada por

$$G_0(x) = e^{-z} \sum_{k=0}^{\infty} \frac{z^k}{k!} x^k = e^{z(x-1)}. \quad (4.38)$$

Fazendo uso da Equação 4.33, obtemos que a função geratriz $G_1(x)$ da rede Erdős-Rényi, para vértices que podem ser alcançados seguindo uma ligação é

$$G_1(x) = \frac{G_0'(x)}{z} = e^{z(x-1)}. \quad (4.39)$$

Das Equações 4.38 e 4.39, verificamos que no caso da distribuição de Poisson $G_1(x) = G_0(x)$. Esta identidade explica porque as propriedades da rede aleatória de Erdős-Rényi são particularmente simples de se resolver analiticamente.

Como segundo exemplo, tomemos a seguinte distribuição de conectividades exponencial

$$P(k) = (1 - e^{-1/\kappa}) e^{-k/\kappa}, \quad (4.40)$$

onde κ é uma constante. Sua função geratriz é dada por

$$G_0(x) = (1 - e^{-1/\kappa}) \sum_{k=0}^{\infty} e^{-k/\kappa} x^k = \frac{1 - e^{-1/\kappa}}{1 - x e^{-1/\kappa}}, \quad (4.41)$$

e de acordo com a Equação 4.33 temos que

$$G_1(x) = \frac{G_0'(x)}{z} = \left[\frac{1 - e^{-1/\kappa}}{1 - x e^{-1/\kappa}} \right]^2. \quad (4.42)$$

onde conectividade média z para essa distribuição de conectividades exponencial é

$$z = \frac{1}{-1 + e^{1/\kappa}}. \quad (4.43)$$

Para o terceiro exemplo, considere um grafo no qual todos os vértices tem conectividade 0, 1, 2, ou 3 com probabilidades p_0, \dots, p_3 . Desta forma, as funções geratrizes assumem as formas do seguintes polinômios

$$G_0(x) = P(3)x^3 + P(2)x^2 + P(1)x + P(0), \quad (4.44)$$

lembrando da Equação 4.33, obtemos

$$G_1(x) = \frac{G'_0(x)}{z} = \frac{3P(3)x^2 + 2P(2)x + P(1)}{3P(3) + 2P(2) + P(1)}, \quad (4.45)$$

e fazendo uso das definições de conectividade média $z = \sum_j jP(j)$ e da distribuição de conectividades excedente $Q(k)$ temos

$$G_1(x) = Q(2)x^2 + Q(1)x + Q(0). \quad (4.46)$$

onde

$$Q(0) = \frac{P(1)}{z}, \quad Q(1) = \frac{2P(2)}{z}, \quad Q(2) = \frac{3P(3)}{z}. \quad (4.47)$$

4.11 Propriedades de redes não-direcionadas

Uma das mais importantes propriedades de redes não-direcionadas e não-conexas é a distribuição de tamanho de componentes na rede. Supondo que a rede esteja num regime abaixo da transição de fase, ou seja, num regime no qual ainda não existe componente gigante, e que o coeficiente de agrupamento seja insignificante. Isso é verdade em se tratando de uma rede aleatória, uma vez a probabilidade de que quaisquer dois vértices distintos i e j escolhidos aleatoriamente, com conectividades k_i e k_j , estejam conectados é a mesma não importando a localização deles na rede. Essa probabilidade é sempre igual a $k_i k_j / (Nz)$, e conseqüentemente tende a zero com $N \rightarrow \infty$, ou seja, nenhuma componente tem laços (*loops*). Em outras palavras, todas as componentes finitas são do tipo árvore (*tree-like*). Essa característica é imprescindível para que a distribuição de tamanho de componentes tenha solução exata.

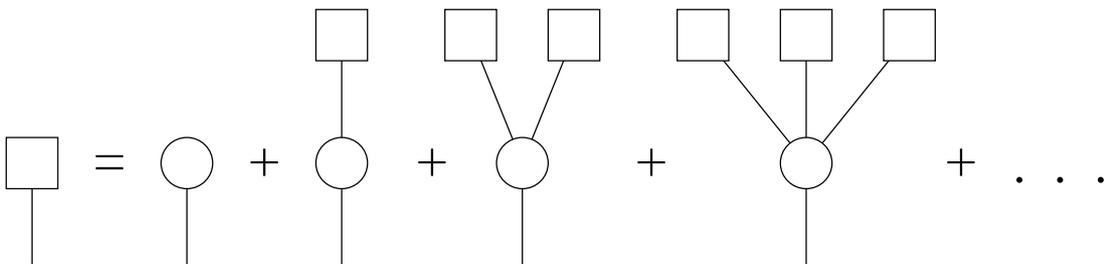


Figura 4.17: Representação das possíveis formas que os componentes de um grafo podem assumir.

Com base nas considerações anteriores, podemos calcular a distribuição de tamanho de componentes, abaixo da transição de fase, da seguinte forma. Considere uma ligação aleatoriamente escolhida. Imagine seguir esta ligação até uma de suas extremidades e a partir desta extremidade alcançar todos os vértices disponíveis através dele. Chamaremos de agrupamento, o conjunto de vértices na extremidade da ligação aleatoriamente escolhida. Seja $H_1(x)$ a função geratriz que gera a distribuição de tamanhos de

agrupamentos de tais agrupamentos, em termo dos números de vértices. Como vemos na Figura 4.17, um agrupamento pode ter várias formas diferentes. Por exemplo, podemos seguir uma ligação aleatoriamente escolhida, e encontrar um único vértice sem nenhuma ligação saindo dele. Uma outra possibilidade é encontrar um vértice com uma ou mais ligação saindo dele. É importante dizer que, cada ligação pode levar a um outro agrupamento completo cujo tamanho também é distribuído de acordo com a função geratriz $H_1(x)$.

Para obter o número total de vértices que podem ser alcançados seguindo por uma ligação aleatoriamente escolhida, precisamos lembrar que o número de ligações k saindo de um vértice, exceto aquele por onde chegamos, é distribuído de acordo com a distribuição $Q(k)$ (veja a Equação 4.24). Outro passo na obtenção de $H_1(x)$ é fazer uso da propriedade multiplicativa da função geratriz, que vimos na Seção 4.10.1. A propriedade multiplicativa afirma que, a distribuição da soma dos tamanhos dos k agrupamentos que podem ser alcançados é gerado por $[H_1(x)]^k$. Então temos que $H_1(x)$ é dado por

$$H_1(x) = x \sum_{k=0}^{\infty} Q(k)[H_1(x)]^k, \quad (4.48)$$

onde o fator x dá conta do vértice na extremidade da ligação aleatoriamente escolhida. Usando a Equação 4.33, podemos reescrever a Equação 4.48 da seguinte forma

$$H_1(x) = xG_1(H_1(x)), \quad (4.49)$$

Contudo, não estamos interessados no número total de vértices que podemos alcançar seguindo por uma ligação aleatoriamente escolhida, mas sim na distribuição de tamanhos dos agrupamentos ao qual pertence um vértice aleatoriamente escolhido. Para calcular

essa distribuição, fazemos uso de que o número de ligações que saem de tal vértice é distribuído de acordo com a distribuição de conectividades $P(k)$, e que cada ligação leva a um agrupamento cujo tamanho, dado em função do número de vértices, é obtido da distribuição gerada pela função $H_1(x)$. Assim, temos que o tamanho da componente completa ao qual pertence um vértice aleatoriamente escolhido é gerado por

$$H_0(x) = x \sum_{k=0}^{\infty} P(k)[H_1(x)]^k = xG_0(H_1(x)). \quad (4.50)$$

que é análogo a probabilidade $P(s)$ de que um vértice aleatoriamente escolhido pertença a um agrupamento de tamanho s

$$H_0(x) = \sum_{s=1}^{\infty} P(s)x^s. \quad (4.51)$$

Tendo as Equações 4.50 e 4.49, podemos calcular a distribuição completa de tamanhos das componentes resolvendo a equação de $H_1(x)$ de maneira auto-consistente e substituindo o resultado na equação de $H_0(x)$. Tomemos a distribuição inicial,

$$H_1(x) = Q(0)x \quad (4.52)$$

podemos encontrar a forma fechada do tamanho da componente completa $H_0(x)$ para qualquer ordem de x , como mostramos a seguir

$$\begin{aligned}
H_1^{(1)}(x) &= Q(0) x \\
H_1^{(2)}(x) &= x \sum_{k=0}^{\infty} Q(k) \left[H_1^{(1)}(x) \right]^k \\
H_1^{(2)}(x) &= Q(0) x + \left[Q(0) Q(1) \right] x^2 \\
H_1^{(3)}(x) &= x \sum_{k=0}^{\infty} Q(k) \left[H_1^{(2)}(x) \right]^k \\
H_1^{(3)}(x) &= Q(0) x + \left[Q(0) Q(1) \right] x^2 \left[Q(0) Q(1)^2 + Q(0)^2 Q(2) \right] x^3 \quad (4.53)
\end{aligned}$$

a Equação 4.53 está correta para a ordem x^3 , mas está incorreta para ordem maior ou igual a x^4 .

Para obtermos as probabilidades $P(s)$ basta desenvolver a Equação 4.51 e comparar seus coeficientes termo a termo com os da Equação 4.53. Abaixo vemos os valores dos $P(s)$, para $s = 1, 2, 3$

$$P(1) = Q(0), \quad P(2) = Q(0) Q(1) \quad \text{e} \quad P(3) = \left[Q(0) Q(1)^2 + Q(0)^2 Q(2) \right]. \quad (4.54)$$

Nem sempre é possível calcular a distribuição de probabilidade de tamanhos das componentes $P(s)$ para todas as ordens de maneira fechada. Contudo, podemos calcular os momentos da distribuição que, em geral, é muito útil. Abaixo da transição de fase, a distribuição de tamanhos das componentes é gerado por $H_0(x)$ (veja Equação 4.50), e conseqüentemente o tamanho médio da componente é

$$\langle s \rangle = H_0'(1) = \left[G_0(H_1(x)) + x G_0'(H_1(x)) H_1'(x) \right]_{x=1} = 1 + G_0'(1) H_1'(1), \quad (4.55)$$

onde usamos a propriedade apresentada na Equação 4.34. O valor de $H'_1(1)$ pode ser calculado da Equação 4.49, diferenciando e rearranjando da seguinte forma

$$H'_1(1) = \frac{1}{1 - G'_1(1)}, \quad (4.56)$$

e substituindo $H'_1(1)$ na Equação 4.55 encontramos que

$$\langle s \rangle = 1 + \frac{G'_0(1)}{1 - G'_1(1)}. \quad (4.57)$$

Esta expressão pode ainda ser escrita como

$$G'_0(1) = \sum_k k P(k) = \langle k \rangle = z_1, \quad (4.58)$$

$$G'_1(1) = \frac{\sum_k k(k-1)P(k)}{\sum_k kP(k)} = \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} = \frac{z_2}{z_1}, \quad (4.59)$$

substituindo $G'_0(1)$ e $G'_1(1)$ na Equação 4.57, obtemos que o tamanho médio das componentes abaixo da transição de fase é

$$\langle s \rangle = 1 + \frac{z_1^2}{z_1 - z_2}. \quad (4.60)$$

Notamos facilmente que esta expressão diverge quando $z_1 = z_2$, o que indica o surgimento de uma componente gigante. Essa é uma derivação rigorosa e mais elegante do que aquela da Seção 4.9.

Os cálculos feitos até agora se referem ao comportamento da rede abaixo da transição de fase, onde não existe componente gigante. Precisamos estender nossos cálculos para o regime acima da transição. A componente gigante, escala com o tamanho

da rede toda, e se torna infinita quando $N \rightarrow \infty$. Isso significa que existirão *loops* na componente gigante, o que faz com que os argumentos usados até agora não são apropriados para calcular a distribuição de tamanho de componentes. Contudo, essa impossibilidade pode ser contornada definindo $H_0(x)$ e $H_1(x)$ como sendo as funções geratrizes para a distribuição de tamanho de componentes, *excluindo a componente gigante*. As componentes não-gigantes são do tipo árvore mesmo no regime acima da transição de fase, assim as Equações 4.49 e 4.50 podem ser utilizadas. A única diferença é que $H_0(1)$ e nem $H_1(1)$ estão mais normalizados a 1 (um). Ao invés disso temos que a distribuição de tamanho de componentes é dada por

$$H_0(1) = \sum_s P(s) = \text{fração de vértices que não pertencem a componente gigante.} \quad (4.61)$$

Este resultado é muito útil, uma vez que nos permite calcular o tamanho S da componente gigante acima da transição de fase como uma fração do tamanho da rede, uma vez que $S = 1 - H_0(1)$. Das Equações 4.49 e 4.50, vemos que S precisa ser a solução das equações

$$S = 1 - G_0(v), \quad v = G_1(v), \quad (4.62)$$

onde $v \equiv H_1(1)$. Assim como no cálculo da distribuição dos tamanhos de componentes para o regime abaixo da transição de fase, as Equações 4.62 normalmente não têm solução fechada. Uma solução pode ser encontrada começando a iteração com um valor inicial de v , por exemplo $v = 0$.

Como no cálculo do tamanho médio das componentes abaixo da transição de fase, também podemos diferenciar a Equação 4.50 para calcular o tamanho médio das componentes não gigantes no regime acima da transição. Diferente do cálculo abaixo da transição, aqui não podemos supor $H_0(1) = H_1(1) = 1$ e teremos que fazer a normalização

de $P(s)$ nós mesmos. Tomando esses cuidados, podemos escrever a expressão do tamanho médio das componentes não-gigante como sendo

$$\begin{aligned}\langle s \rangle &= \frac{H'_0(1)}{H_0(1)} = \frac{1}{H_0(1)} \left[G_0(H_1(1)) + \frac{G'_0(H_1(1))G_1(H_1(1))}{1 - G'_1(H_1(1))} \right] \\ &= 1 + \frac{z\nu^2}{[1 - S][1 - G'_1(\nu)]},\end{aligned}\quad (4.63)$$

onde ν e S são obtidos da Equação 4.62. Podemos re-obter a equação do tamanho médio das componentes (veja Equação 4.57), fazendo $S = 0$, $\nu = 1$.

4.12 Watts e Strogatz e as Redes Mundo Pequeno

Contrário da rede ER, muitas redes reais possuem coeficiente de agrupamento alto (veja Tabela 4.2). Watts e Strogatz, na tentativa de unir a características de mundo pequeno presente e de coeficiente de agrupamento alto numa só rede, propuseram a construção de uma rede, o modelo Watts-Strogatz **WS**, que fosse inicialmente uma rede regular e que então fossem adicionadas ou movidas ligações, com o objetivo criar “atalhos” entre pontos distantes da rede.

| Tipo de Rede | N | m | z | ℓ | C | Ref. |
|--------------------------|---------|----------|--------|--------|-------|---------|
| atores de filmes | 449913 | 25516482 | 113,43 | 3,48 | 0,20 | [31,32] |
| diretores de companhias | 7673 | 55392 | 14,44 | 4,60 | 0,59 | [91,93] |
| colaboração (Matemática) | 253339 | 496489 | 3,92 | 7,57 | 0,15 | [94,95] |
| colaboração (Física) | 52909 | 245300 | 9,27 | 6,19 | 0,45 | [25,26] |
| colaboração (Biologia) | 1520251 | 11803064 | 15,53 | 4,92 | 0,088 | [25,26] |
| jogadores de futebol | 13411 | 315566 | 47,10 | 2,84 | 0,790 | [25,26] |
| Erdős-Rényi | 13411 | 315443 | 47,10 | 2,84 | 0,004 | [25,26] |

Tabela 4.2: As propriedade medidas são: número total de vértices N ; número total de ligações m ; conectividade média z ; distância média entre vértices ℓ ; coeficiente de agrupamento C . Na última coluna estão algumas referências de onde os dados foram obtidos.

O modelo **WS** pode ser construído em redes de qualquer dimensão ou topologia, mas foi melhor estudada para no caso da rede unidimensional, sendo definido da seguinte maneira: tome uma rede unidimensional de L vértices com condição periódica de contorno, ou seja, um anel. Então, conecte cada vértice a seus primeiros k vizinhos (vértices que estejam o mais próximos do vértice em questão). Assim obtemos uma rede com Lk ligações (veja exemplo de rede com $L = 16$ e $k = 2$). O modelo Watts-Strogatz é então criado “reconectando” uma fração das ligações dessa rede. O processo de reconexão consiste em visitar cada ligação da rede e, com probabilidade p , reconectar uma das extremidades da ligação a um novo vértice escolhido aleatoriamente na rede. No modelo **WS** são proibidas as formações de ligações duplas e de auto-ligações. Esse processo é ilustrado na Figura 4.18.

O processo de reconexão permite ao modelo **WS** transformar uma rede com característica de rede regular em uma rede aleatória. Quando $p = 0$, temos uma rede regular, e o coeficiente de agrupamento dessa rede vale

$$C = \frac{3k - 3}{4k - 2}, \quad (4.64)$$

que tende a $3/4$ quanto $k \rightarrow \infty$. Contudo, a rede regular não tem o efeito mundo pequeno. Quando $p = 1$, cada ligação é reconectada a um novo vértice transformando a rede regular numa rede aleatória, com distância geodésica típica da ordem de $\ln L / \ln k$, mas com coeficiente de agrupamento muito baixo $C \simeq 2k/L$. Contudo, como Watts e Strogatz mostraram, em cálculos numéricos do coeficiente de agrupamento e da distância geodésica, existe um razoável intervalo da probabilidade p para o qual o modelo apresenta simultaneamente distância geodésica baixa e coeficiente de agrupamento alto — veja Figura 4.19.

O modelo **WS** teve como inspiração sistemas sociais, basta lembrar que muitas pessoas são amigas de seus vizinhos — vizinhos da mesma rua, colegas de trabalho, pessoas que seus amigos apresentam a você. Todos nós temos ainda, pelo menos

um amigo que está longe. Esses relacionamentos de longa distância são representados pelas ligações de longo alcance obtidas pelas reconexões no modelo **WS**.

Para entender, um pouco mais, a existência de curtos caminhos entre dois vértices quaisquer e agrupamento, se faz necessário analisar o comportamento do coeficiente de agrupamento $C(p)$ e o comprimento médio dos menores caminhos $\ell(p)$ em função da probabilidade de reconexão p .

Para uma rede com condições periódicas de contorno, fazendo $p = 0$ temos que $\ell(0) \simeq L/2k \gg 1$ e $C(0) \simeq 3/4$. Vemos então que ℓ escala linearmente com o tamanho da rede e o C é alto, assim como deve ser em uma rede regular. Quando $p \rightarrow 1$ o modelo SW converge para uma rede aleatória, onde temos $\ell(1) \sim \ln(L)/\ln(k)$ e $C(1) \sim k/L$. Vemos que ℓ escala com o logaritmo do tamanho da rede, enquanto o coeficiente de agrupamento escala com o inverso do tamanho da rede.

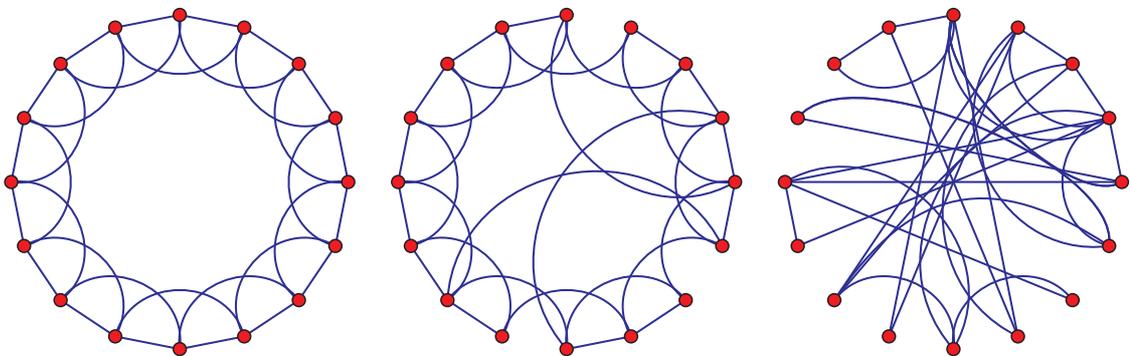


Figura 4.18: Procedimento de reconectar os vértices do modelo WS que transforma uma rede (circular) regular em uma rede aleatória sem fazer mudança no número de vértice ou de ligação. Começamos com uma rede de $L = 16$ vértices, cada um conectado aos seus 4 vizinhos mais próximos, em seguida escolhemos um vértice e uma ligação que conecta um de seus vizinhos. Com uma probabilidade p reconectamos essa ligação a um outro vértice escolhido aleatoriamente, fazemos isso até que todas as ligações, da rede original, tenham sido considerados. Realizações deste procedimento é mostrado, aqui nesta figura, para três valores diferentes de p . Na primeira, foi usado $p = 0$, com isso a rede original permanece inalterada. Na segunda, com um valor de p no intervalo $0 \leq p \leq 1$. Na terceira, é obtida uma rede totalmente aleatória com $p = 1$, onde todas as ligações foram aleatoriamente reconectadas.

A análise desses casos limites pode nos levar a concluir que um valor de coeficiente de agrupamento $C(p)$ alto está sempre associado a um valor do comprimento médio dos menores caminhos $\ell(p)$ pequeno. Contudo, como vemos na Figura 4.19, existe

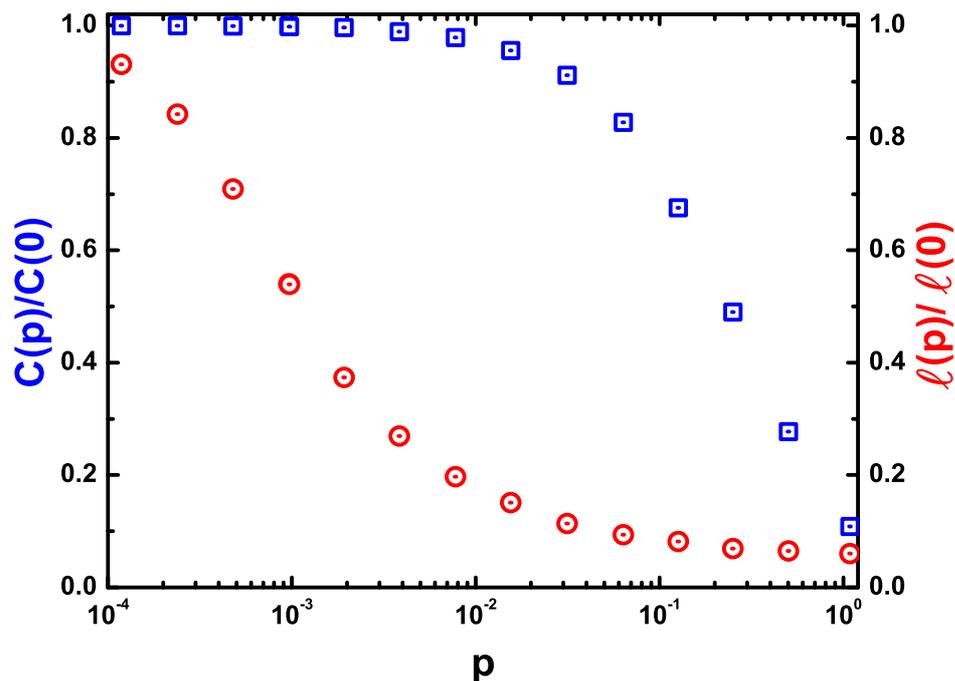


Figura 4.19: Coeficiente de agrupamento $C(p)$ e o comprimento médio dos menores caminhos $\ell(p)$ do modelo Watts-Strogatz. Os dados foram normalizados pelos valores de $C(0)$ e $\ell(0)$, respectivamente. Pode ser visto a rápida queda de $\ell(p)$, demonstrando que a rede se comporta como uma rede mundo pequeno. Durante essa queda de $\ell(p)$, $C(p)$ permanece praticamente constante, o que indica que a transição para uma rede mundo pequeno é quase imperceptível localmente.

um considerável intervalo de valores de p onde $\ell(p)$ tem valores próximos de $\ell(1)$ e o coeficiente de agrupamento ainda é bem maior do que $C(1)$. A existência desse intervalo de valores de p tem sua origem na rápida queda de $\ell(p)$ para pequenos valores de p enquanto $C(p)$ permanece quase que inalterado. Nesse regime a rede possui valores altos para o coeficiente de agrupamento e valores baixos para o comprimento médio dos menores caminhos.

4.13 Barabási e Albert e as Redes Livres de Escala

Em 1982, um dos alunos de Paul Erdős, chamado Béla Bollobás, provou exatamente a forma da distribuição de conectividades do modelo Erdős-Rényi. O resultado mostra que a distribuição é de Poisson. Essa distribuição tem algumas características marcantes, seu pico é pronunciado, indicando que a maioria dos vértices têm, em média, o mesmo número de ligações. Nos extremos da distribuição de Poisson estão representados os vértices raros, com conectividade muito baixa (à esquerda do pico) ou com conectividade muito alta (à direita do pico).

Quando Albert-László Barabási, em 1998, instruiu seu colaborador Hawoong Jeong a construir um robô¹⁴ para mapear a rede WWW (páginas na internet), eles tinham a expectativa de que a rede que iria se revelar por trás da WWW possuísse características de rede aleatória. Assim sendo, eles esperavam observar uma distribuição com um pico, evidenciando que em média as páginas fossem igualmente populares. Mas quando o robô finalizou a tarefa, os dados finais do levantamento causaram espanto. Os dados indicavam que a maioria dos vértices tinham poucas ligações e que alguns poucos vértices concentravam a maior parte das ligações. Quando tentaram ajustar o histograma de conectividades dos vértices, tiveram uma grande surpresa. Visto que a distribuição de Poisson não servia para ajustar o conjunto de pontos, eles tentaram outras distribuições e eis que quando os dados foram dispostos em escala log-log, eles se ajustaram muito bem. O conjunto de pontos decaía como uma lei de potência. A distribuição de conectividades das páginas que recebiam ligações de outras páginas, seguia uma lei de potência com um expoente $\gamma_{in} \approx 2,1$. Também foi medida a distribuição de conectividades das páginas que enviam ligações para outras páginas e encontraram uma lei de potência com um expoente $\gamma_{out} \approx 2,5$.

¹⁴A primeira tarefa do robô, construído por Jeong, foi mapear as 300 mil páginas do domínio nd.edu (Universidade de Notre Dame). O objetivo era fazer com o que o robô viajasse pelas ligações existentes entre as páginas (não importando o seu conteúdo).

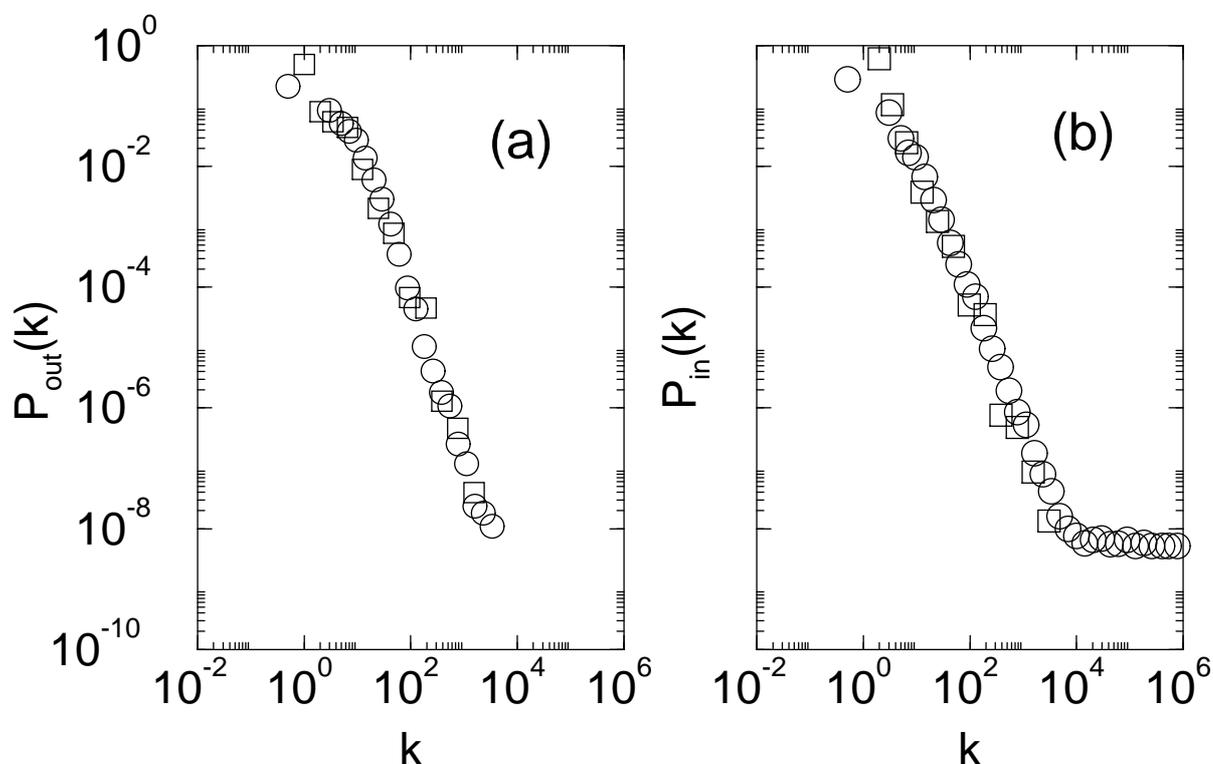


Figura 4.20: Distribuição de conectividades de redes World-Wide Web de diferentes tamanhos. Os quadrados correspondem a uma amostra de 325.729 de páginas (obtidos por Réka Albert) e os círculos correspondem a uma medida de 200 milhões de páginas (obtidos por Broder et al.). (a) Distribuição de conectividades de ligações que recebem ligações de outras páginas. (b) Distribuição de conectividades de ligações que enviam ligações para outras páginas.

Depois do estudo da rede WWW, o grupo de Barabási também analisou a rede dos filmes de Hollywood, onde dois atores estavam ligados se eles participaram de um mesmo filme, e verificou que o número de atores que tinha exatamente k ligações com outros atores decaía seguindo novamente uma lei de potência. Depois descobriram um trabalho de Sidney Redner [96,97], sobre a distribuição de citações em revistas de física que também seguia uma lei de potência. Em seguida, começaram a surgir trabalhos e mais trabalhos estudando redes e em muitos deles uma característica recorrente: o número de vértices, da rede em questão, com exatamente k ligações seguia uma lei de potência. Cada novo sistema estudado verificava-se que cada um possuía um único expoente e entre os

sistemas estudados os valores dos expoentes das distribuições estavam sempre entre dois e três.

Há algumas décadas o meio científico tem verificado que em algumas ocasiões a natureza gera quantidades que são regidas por uma distribuição lei de potência, indicando a possibilidade de coexistência de eventos de todas as escalas. Assim, a importância das leis de potências, não reside simplesmente no fato dela mostrar a existência de eventos de menores ou maiores escalas, mas sim por tornar evidente a coexistência de eventos de todas as escalas. Lembrando que, eventos de escalas muito pequenas e muito grandes são "proibidos" na distribuição de Poisson. As distribuições do tipo lei de potência são muito diferentes das curvas do tipo sino (veja as Figuras 4.20 e 4.21).

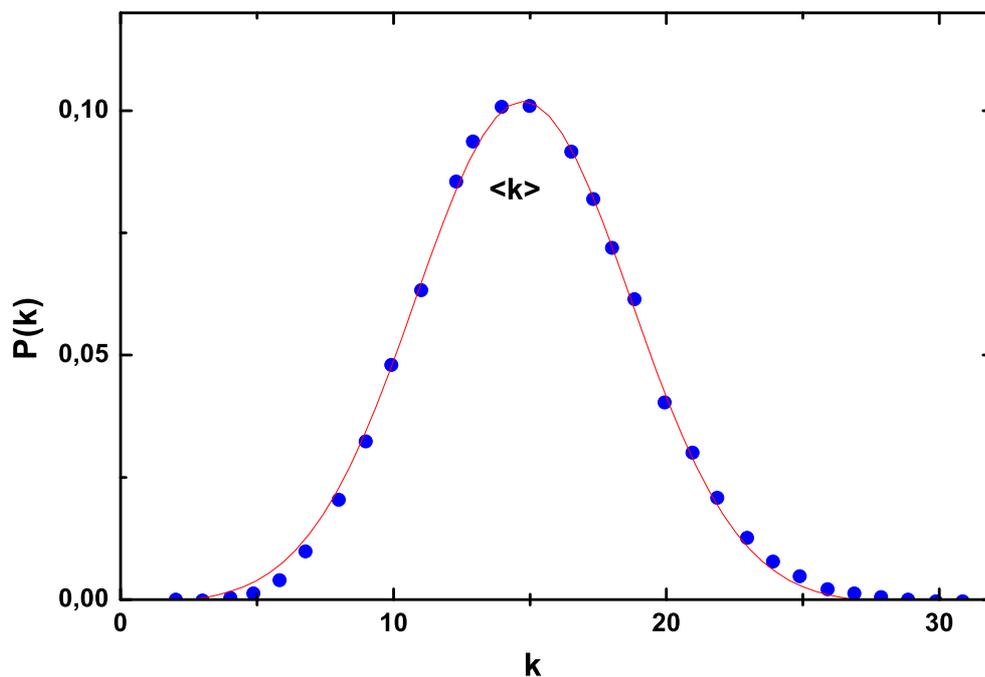


Figura 4.21: A distribuição de conectividades de uma rede aleatória, com $N = 10.000$ vértices e probabilidade de conexão $p = 0,0015$. A linha contínua mostra o ajuste da distribuição de conectividades segundo uma distribuição de Poisson.

Sistemas como a sociedade, as células que compõe nosso corpo e tantos outros são mapeados em redes complexas, mas suas ligações não são aleatórias. Uma interação social, entre duas pessoas ou mais pessoas, não é totalmente casual, afinal de contas você não planeja uma festa de aniversário e sai pela rua chamando qualquer pessoa que você encontre para participar da sua festa, ou ainda um gerente de banco não concede um empréstimo aos seus correntistas com base no resultado de um jogo de dados.

Muitos de nós cremos ou queremos acreditar que não vivemos num mundo fortuito e com base nessa crença temos esperança de que exista uma certa ordem nos sistemas complexos, do qual também faz parte nosso mundo. Parece ser natural que façamos a seguinte pergunta: como puderam, dois matemáticos prestigiados como Erdős e Rényi, modelar a complexidade das redes como sendo um processo totalmente aleatório como é um grafo aleatório? A resposta é mais simples do que a pergunta. Aparentemente, eles nunca tiveram em mente criar uma teoria universal de formação de redes. Eles, como bons matemáticos, estavam mais interessados pelas belezas matemáticas, que os grafos aleatórios possuem. Para eles o fato do modelo Erdős-Rényi descrever ou não sistemas reais, era secundário.

Muitos sistemas reais não podem ser tratados como redes aleatórias, uma forte evidência está no fato que a distribuição de conectividades (ou graus) de muitas redes seguem uma lei de potência, $P(k) \sim k^{-\gamma}$. Uma vez confirmado que grafos aleatórios nem sempre podem ser utilizados no estudo de redes reais, visto que não capturam suas características, como por exemplo: comportamento de escala livre da distribuição de conectividades, se faz necessário estudar novos modelos que melhor descrevam as redes reais. E essa é nossa motivação para no Capítulo 5 estudar o modelo Barabási-Albert não-linear.

CAPÍTULO 5

Rede Barabási-Albert Não-Linear

Nos últimos anos, a comunidade científica passou a se interessar pela estrutura e a dinâmica das redes complexas. Como vimos no Capítulo 4, isso se deve à convergência de interesses entre diferentes áreas tais como física, biologia, sociologia que transformaram o estudo de redes complexas numa área interdisciplinar.

Uma característica comum aos modelos **ER** e **WS**, apresentados no Capítulo 4, é que ambos têm suas distribuições de conectividades $P(k)$ decaindo exponencialmente, e no caso do modelo **ER** a conectividade média $\langle k \rangle$ é uma característica que depende de uma probabilidade p . Contudo, muitos sistemas na natureza parecem ter suas distribuições de conectividades $P(k)$ decaindo como lei de potência. Para entender a origem dessa divergência de comportamento, Barabási e Albert argumentaram que existe dois aspectos genéricos das redes reais que não estão incorporados nos modelos **ER** e **WS**. O primeiro aspecto é que os dois modelos trabalham com um número fixo N de vértices, que são aleatoriamente conectados (no caso do modelo **ER**) ou reconectados (no caso do modelo **WS**). Entretanto, muitas redes reais crescem continuamente, ou seja, o número de vértices N da rede cresce com o tempo. O segundo aspecto é que ambos os modelos assumem que a probabilidade de que dois vértices estejam conectados é aleatória. Contudo, muitas redes reais exibem a característica de ligação preferencial de seus vértices. Por exemplo, é mais provável uma nova página na internet tente se ligar a páginas bem conhecidas, assim como é mais provável um artigo citar os artigos mais conhecidos da área.

Estes exemplos indicam que a probabilidade de que um novo vértice se conecte a um outro vértice existente na rede não é aleatória, ao contrário existe uma maior probabilidade de um novo vértice se conectar a um vértice já existente que tenha um grande número de ligações. O modelo de escala livre proposto por Barabási e Albert (**BA**), incorpora esses dois aspectos. O modelo é definido em dois passos:

- i) Crescimento: começamos com um número pequeno (m) de vértices, conectados entre si e a cada passo temporal t adicionamos um novo vértice com m ligações que se conectarão aos vértices já existentes.
- ii) Ligação preferencial: escolhemos os vértices com os quais o novo vértice vai se conectar, levando em conta que a probabilidade Π de que o novo vértice se conecte ao i -ésimo vértice depende da sua conectividade k_i da seguinte forma:

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}, \quad (5.1)$$

Depois de t passos temporais a rede produzida tem $N = t + m$ vértices e $m t$ ligações. E constatamos que essa rede evolui num estado de escala livre, onde a probabilidade de que um vértice tenha k ligações segue uma lei de potência com um expoente $\gamma = 2,9 \pm 0,1$, veja Figura 5. Esse expoente independe de m , único parâmetro do modelo **BA**.

O modelo de rede de escala livre mais estudado é o Barabási-Albert (**BA**) linear [14], que acabamos de apresentar. Na sua forma generalizada, que chamamos de Barabási-Albert não-linear, a rede é construída da seguinte forma: cada novo vértice a ser adicionado a rede, se conecta a outro vértice i (já existente) com probabilidade

$$\Pi(k_i) = \frac{k_i^\alpha}{\sum_j k_j^\alpha}, \quad (5.2)$$

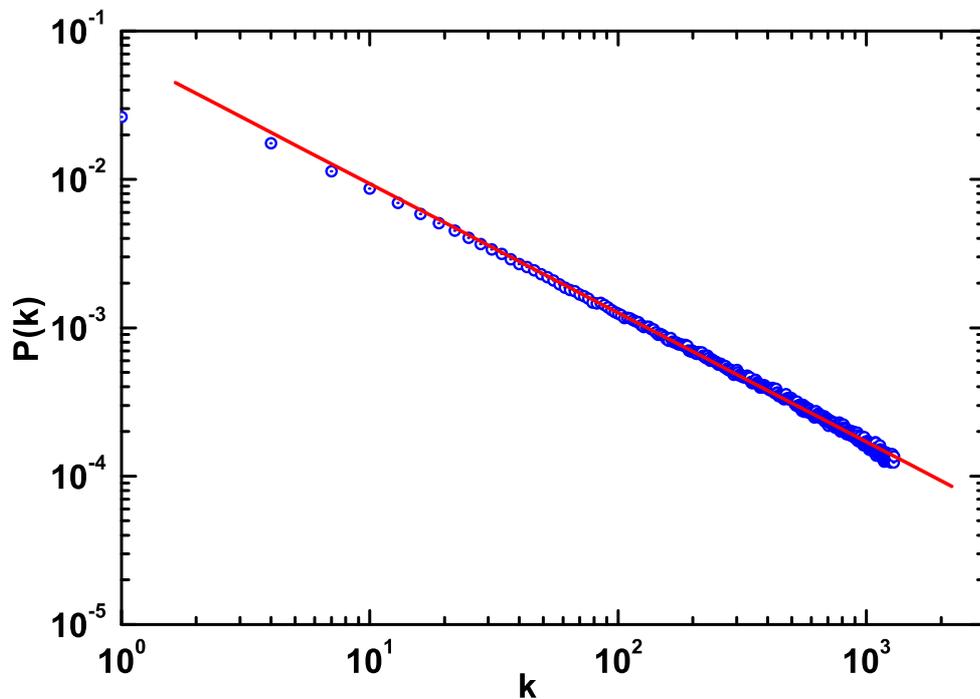


Figura 5.1: Distribuição de conectividades do modelo **BA**, com $N = m + t = 300\,000$ e $m = 2$).

onde k_i é a conectividade, ou seja, o número de ligações de cada vértice i e α é um número real. Cada novo vértice adicionado contribui com m novas ligações à rede. No caso linear ($\alpha = 1$), a rede tem algumas características desejáveis tais como ser de escala livre, ter efeito mundo pequeno, mas também possui alguns outros aspectos que pouco desejáveis tais como, não ser assortativo¹ e o coeficiente de agrupamento ir a zero com o aumento do tamanho da rede. Como já vimos no Capítulo 4, uma rede é dita assortativa se seus vértices se conectam a vértices de mesma conectividade, ou seja, vértices de pequena conectividade se conectam a outros vértices de pequena conectividade e vértices de grande conectividade se conectam a outros vértices de grande conectividade e assim por diante.

¹Assortativo: relativo a preferência de se ligar preferencialmente a indivíduos iguais ou com características semelhantes

Quando pensamos em redes sociais, características como assortatividade e coeficiente de agrupamento alto são desejadas.

Neste Capítulo estudamos o modelo **BA** não-linear determinando algumas importantes quantidades tais como distribuição conectividade $P(k)$, coeficiente de agrupamento C , coeficiente de assortatividade r e a comprimento médio do menor caminho l . O estudo foi feito sobre todo o espaço de parâmetros m e α .

5.1 A Distribuição de Conectividades Através da Equação Mestra

A distribuição de conectividades $P(k)$, de que um vértice aleatoriamente escolhido tenha conectividade k já foi obtido para o caso de $m = 1$ [96, 97]. Aqui nesta Seção, apresentamos o cálculo de $P(k)$ para qualquer valor de m . Para uma rede com N vértices, a equação mestra pode ser escrita como

$$\frac{dN(k, t)}{dt} = \frac{m}{M_\alpha(t)} [(k-1)^\alpha N(k-1, t) - k^\alpha N(k, t)] + \delta_{k,m}, \quad \text{com } k \geq m \text{ e } t \geq 1, \quad (5.3)$$

onde $N(k, t)$ é o número de vértices com conectividade k no instante de tempo t ,

$$M_\alpha(t) = \sum_{k=m}^N k^\alpha N(k, t), \quad t \geq 1, \quad (5.4)$$

é a normalização adequada. O primeiro termo do lado direito da Equação 5.3 leva em conta o processo no qual o novo vértice se conecta a um outro com $k-1$ ligações, tornando-o, assim, um vértice de conectividade k . O segundo termo dessa equação leva em conta

o processo no qual o novo vértice se conecta a um outro com k ligações, tornando-o, assim, um vértice de $k + 1$ ligações. Para cada uma das m ligações do novo vértice, isto acontece com probabilidade $m(k - 1)^\alpha / M_\alpha(t)$ (no caso do primeiro termo da equação) e $mk^\alpha / M_\alpha(t)$ (no caso do segundo termo da equação). O último termo considera que um novo vértice de conectividade m foi adicionado à rede.

Podemos obter o n -ésimo momento $M_n(t)$ ($\alpha = n =$ número inteiro na Equação 5.4, onde $M_0(0) = m - 1$) da distribuição de conectividades. No caso do momento de ordem zero ($M_0(t)$), basta fazer $\alpha = 0$ na Equação 5.3 e em seguida somar sobre todos os k 's, obtendo

$$\frac{d}{dt} \left(\sum_{k=m}^{\infty} N(k, t) \right) = \sum_{k=m}^{\infty} \left(\frac{m}{M_0(t)} [N(k-1, t) - N(k, t)] + \delta_{k,m} \right), \quad (5.5)$$

mas de 5.4, temos

$$M_0(t) = \sum_{k=m}^{\infty} N(k, t), \quad (5.6)$$

e reescrevendo a Equação 5.5 temos

$$\begin{aligned} \frac{d}{dt} \left(M_0(t) \right) &= \frac{m}{M_0(t)} [M_0 - M_0] + 1, \\ \frac{d}{dt} \left(M_0(t) \right) &= 1, \end{aligned} \quad (5.7)$$

integrando a Equação 5.7 temos que

$$M_0(t) = M_0(0) + t. \quad (5.8)$$

Para obter o momento seguinte ($M_1(t)$), basta fazer $\alpha = 1$ e multiplicar a Equação 5.3 por k e somar em k

$$\frac{d}{dt} \left(\sum_{k=m}^{\infty} k N(k, t) \right) = \sum_{k=m}^{\infty} \left(\frac{m}{M_1(t)} [k(k-1)N(k-1, t) - k k N(k, t)] + k \delta_{k,m} \right), \quad (5.9)$$

lembrando que $M_1(t) = \sum_k k N(k, t)$, podemos reescrever a Equação 5.9 da seguinte forma

$$\frac{d}{dt} \left(M_1(t) \right) = \frac{m}{M_1(t)} \sum_{k=m}^{\infty} \left[k(k-1)N(k-1, t) - k k N(k, t) \right] + \sum_{k=m}^{\infty} k \delta_{k,m}, \quad (5.10)$$

e desenvolvendo a soma em k temos que

$$\begin{aligned} \frac{d}{dt} \left(M_1(t) \right) = \frac{m}{M_1(t)} & \left[(m)(m-1)N(m-1, t) - (m)(m)N(m, t) + \right. \\ & (m+1)(m)N(m, t) - (m+1)(m+1)N(m+1, t) + \\ & (m+2)(m+1)N(m+1, t) - (m+2)(m+2)N(m+2, t) + \\ & (m+3)(m+2)N(m+2, t) - (m+3)(m+3)N(m+3, t) + \\ & \left. (m+4)(m+3)N(m+3, t) + \dots \right] + m, \quad (5.11) \end{aligned}$$

onde $N(m - 1, t) = 0$ pois não existe nenhum vértice na rede com conectividade $m - 1$, rearranjando a Equação 5.11 e fazendo uso da definição de M_1 temos que

$$\begin{aligned} \frac{d}{dt} \left(M_1(t) \right) &= \frac{m}{M_1(t)} \left[(m)N(m, t) + (m + 1)N(m + 1, t) + (m + 2)N(m + 2, t) + \right. \\ &\quad \left. (m + 3)N(m + 3, t) + \dots \right] + m \\ \frac{d}{dt} \left(M_1(t) \right) &= \frac{m}{M_1(t)} \left[M_1(t) \right] + m, \end{aligned} \quad (5.12)$$

e então, integrando a Equação 5.12 obtemos

$$M_1(t) = M_1(0) + 2 m t. \quad (5.13)$$

Para o caso linear, $\alpha = 1$ e $m = 1$, a Equação 5.3 pode ser resolvida para uma condição inicial arbitrária. No regime de $t \rightarrow \infty$, onde as condições iniciais são irrelevantes, podemos escrever $M_1(t) = 2 t$ e da Equação 5.3 podemos calcular o número de vértice com conectividade 1, isto é $N(1, t)$ fazendo

$$\frac{d}{dt} N(1, t) = \frac{1}{M_1(t)} \left[0 - N(1, t) \right] + 1, \quad (5.14)$$

integrando a Equação 5.14, temos que

$$\begin{aligned} N(1, t) &= -\frac{N(1, t)}{2} + t \\ N(1, t) &= \frac{2}{3} t. \end{aligned} \quad (5.15)$$

Usando a Equação 5.3 podemos obter também o número de vértices com conectividade 2, $N(2, t)$ fazendo

$$\frac{d}{dt}N(2, t) = \frac{1}{M_1(t)} \left[N(1, t) - 2N(2, t) \right] + 1 \quad (5.16)$$

integrando 5.16 e usando o resultado de 5.15, temos que o número de vértice com conectividade 2 é dado por

$$\begin{aligned} N(2, t) &= \frac{N(1, t) - 2N(2, t)}{2} + t \\ N(2, t) &= \frac{t}{6} \end{aligned} \quad (5.17)$$

Os resultados das equações anteriores, sugerem um crescimento linear de $N(k, t)$ com o tempo. Algo coerente, uma vez que

$$P(k, t) = \frac{N(k, t)}{N} \sim \frac{N(k, t)}{t}, \quad (5.18)$$

ou seja, $N(k, t)$ é uma distribuição estacionária para $P(k, t)$, onde $P(k, t)$ é a probabilidade de que um vértice tenha k ligações no instante de tempo t . Podemos reescrever 5.18 como $N(k, t) = t P(k, t)$, e então substituindo $N(k, t)$ na Equação 5.3 obtemos a relação recursiva

$$P(k, t) = P(k-1, t) \frac{(k-1)}{(k+2)}, \quad (5.19)$$

cuja solução é

$$P(k, t) = \frac{4}{k(k+1)(k+2)} \quad (5.20)$$

No caso geral em que $-\infty \leq \alpha \leq 1$, a distribuição de conectividades $P(k, t)$ gera um $M_\alpha(t)$ que tem a seguinte forma

$$M_\alpha(t) = \mu t, \quad 0 \leq \mu \leq 2. \quad (5.21)$$

Fazendo uso da Equação 5.3, com $k = m$, temos que

$$\frac{d}{dt} N(m, t) = -\frac{m^{1+\alpha}}{M_\alpha(t)} N(m, t) + 1, \quad (5.22)$$

lembrando de 5.18 e 5.21 podemos reescrever a Equação 5.22 da seguinte forma

$$\frac{d}{dt} \left[t P(m, t) \right] = -\frac{m^{1+\alpha}}{\mu} P(m, t) + 1, \quad (5.23)$$

e derivando o lado esquerdo da Equação 5.23, levando em conta que $P(m, t)$ está no estado estacionário, temos que

$$P(m, t) = -\frac{m^{1+\alpha}}{\mu} P(m, t) + 1, \quad (5.24)$$

e isolando $P(m, t)$ finalmente obtemos que

$$P(m, t) = \frac{\mu}{\mu + m^{1+\alpha}}. \quad (5.25)$$

Assim como no caso de $k = m$, também podemos resolver a Equação 5.3 para $k \geq m + 1$ obtendo

$$P(k, t) = \frac{m(k-1)^\alpha}{\mu + m k^\alpha} P(k-1, t), \quad k \geq m + 1, \quad (5.26)$$

que tem como solução a seguinte equação,

$$P(k, t) = \frac{\mu}{m k^\alpha} \prod_{j=m}^k \left(1 + \frac{\mu}{m j^\alpha}\right)^{-1}, \quad (5.27)$$

e como vemos na Figura 5.2 a Equação (5.27) tem boa concordância com resultados obti-

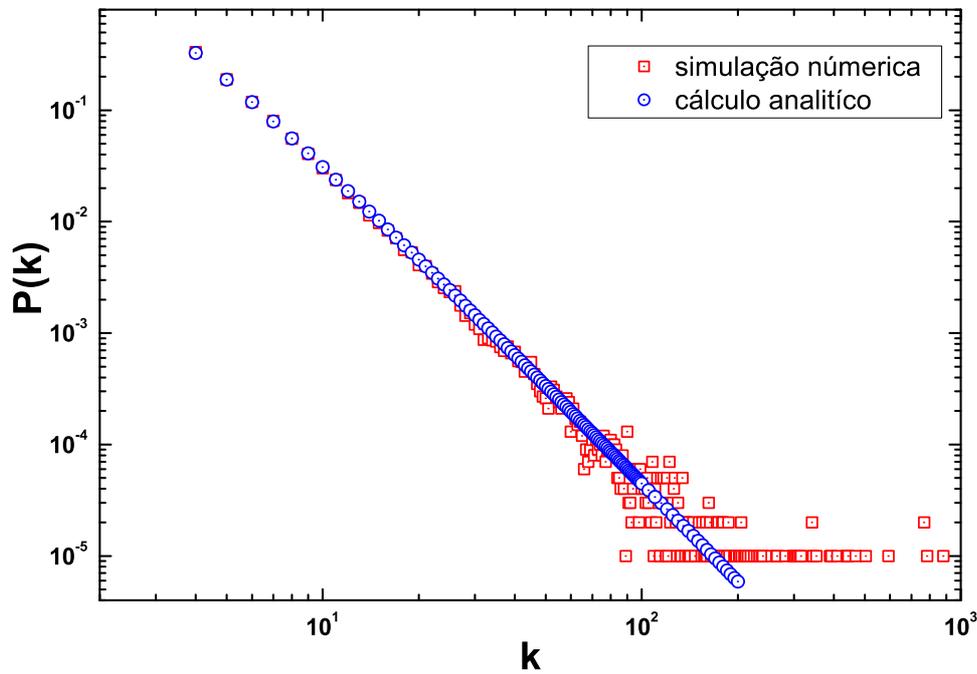


Figura 5.2: Distribuição de conectividades de uma rede BA-não linear para o caso de $\alpha = 1$ e $m = 4$. Os quadrados vermelhos são pontos que foram obtidos via simulação numérica, enquanto os círculos azuis foram obtidos via cálculo analítico Eq. (5.27)

dos numericamente, via simulação.

Podemos agora estabelecer a dependência da amplitude μ com relação a α .

Substituindo $N(k, t) = t P(k, t)$ na Equação 5.4 obtemos

$$M_\alpha(t) = \sum_{k=m}^N k^\alpha t P(k, t), \quad (5.28)$$

e lembrando da Equação 5.21 que $\mu = M_\alpha(t)/t$, podemos escrever μ como

$$\mu = \sum_{k \geq m} k^\alpha P(k, t) \quad (5.29)$$

junto com a Equação 5.27 obtemos a relação intrínseca

$$\mu = m^\alpha \left[1 - m + \sum_{k=m+1}^{\infty} \prod_{j=m+1}^k \left(1 + \frac{\mu}{m j^\alpha} \right)^{-1} \right]. \quad (5.30)$$

Não é fácil extrair informações explícitas sobre a distribuição de conectividades da rede. Duas exceções são os casos limites $\alpha = 0$ e $\alpha = 1$, para os quais $\mu = 1$ e $\mu = 2$, respectivamente. As duas distribuições de conectividades são dadas por $P(k) = 2^{-k}$ (no caso de $m=1$) e $P(k) \sim k^{-3}$ (veja Equação 5.20). Na Figura 5.1 vemos o comportamento de μ em função de α , no intervalo de $-6 \leq \alpha \leq 1$, obtido da resolução numérica (com o auxílio do Maple) da Equação 5.30.

Quando $\alpha < 0$, as expressões obtidas até o momento ainda são válidas mas, o modelo **BA** não-linear perde sua característica de ligação preferencial, o que podemos interpretar como a perda da característica conhecida como *rich get richer* (os ricos tornam-se cada vez mais ricos). Ou seja, quando α é negativo, cada novo vértice adicionado na rede se conecta a vértices de baixa conectividade, algo que podemos interpretar como uma “tendência socialista”. No caso de valores de α muito negativos a tendência é de que no final da construção da rede todos os vértices tenham em média a mesma conectividade k [98]. Senão vejamos, escrevendo $\alpha = -|\alpha|$, assumindo $|\alpha| \gg 1$ e a conectividade k como sendo uma variável contínua, a Equação 5.27 de distribuição de conectividades pode ser escrita como

$$P(k, t) = \frac{\mu}{m} e^{|\alpha| \ln k - \int_m^k \ln \left(1 + \frac{\mu}{m} x^{|\alpha|} \right) dx}, \quad (5.31)$$

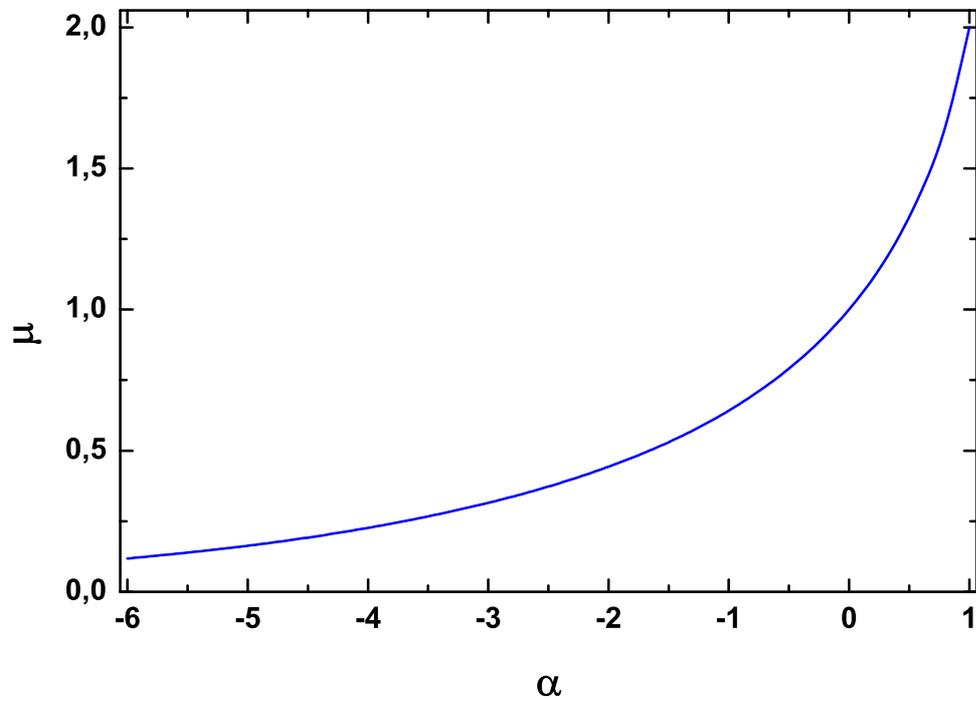


Figura 5.3: Comportamento de μ em função de α .

como $0 \leq P(k, t) \leq 1$, temos que

$$\lim_{|\alpha| \rightarrow \infty} \mu = \sum_m^{\infty} k^{-|\alpha|} P(k, t) \rightarrow 0. \quad (5.32)$$

então podemos reescrever a Equação 5.31 da seguinte forma

$$P(k, t) = \frac{\mu}{m} e^{|\alpha| \ln k - \int_m^k \frac{\mu}{m} x^{|\alpha|} dx}, \quad (5.33)$$

pois $\mu \rightarrow 0$, e com isso

$$\ln\left(1 + \frac{\mu}{m}x^{|\alpha|}\right) \sim \frac{\mu}{m}x^{|\alpha|}. \quad (5.34)$$

Resolvendo a Equação 5.33 de distribuição de conectividades e tomando a primeira e segunda derivada de $P(k, t)$ em k , verificamos que essa distribuição de conectividades tem um valor máximo em

$$k_{max} = \left[\frac{m|\alpha|}{\mu} \right]^{\frac{1}{|\alpha|}}. \quad (5.35)$$

A distribuição de conectividades expandida em volta deste valor máximo tem a forma gaussiana

$$P(k, t) \sim |\alpha| \exp\left\{-\frac{|\alpha|^2}{2}(k - k_{max})^2\right\}. \quad (5.36)$$

que no limite $|\alpha| \rightarrow \infty$ se aproxima da função delta de Dirac $\delta(k - k_{max})$, onde

$$k_{max} = \langle k \rangle = \sum_k k P(k, t) = 2m \quad (5.37)$$

Com isso, a rede se torna homogênea, ou seja, a rede tem todos os seus vértices com a mesma conectividade $2m$.

No caso de $\alpha > 1$, o comportamento assintótico de M_α é t^α [97]. Nesse regime, o modelo exhibe o fenômeno *winner takes all* (o vencedor leva tudo). Quando construímos uma rede usando $\alpha > 1$ e $m = 1$ surge um único vértice super conectado que está ligado com quase todos os outros vértices. Para $m > 1$, encontramos m vértices super conectados. Este resultado foi confirmado exhaustivamente por nossas simulações numéricas. Para redes suficientemente grandes, o histograma de conectividades apresenta sempre m vértices super conectados.

Um comportamento singular acontece quando $\alpha > 2$, nesses casos, existe uma probabilidade B , não nula, de que os vértices iniciais (totalmente conexos) estejam

ligados a todos os outros vértices da rede. Para determinar essa probabilidade vamos considerar um processo de tempo discreto onde um vértice é introduzido a cada passo temporal e este sempre se conecta aos vértices iniciais. Fazendo $s = t - 1$ ($t = N - m$), quando o $(m + 1)$ -ésimo vértice estiver sendo adicionado, a probabilidade de que este vértice se conecte aos m vértices iniciais é dado por

$$\theta(s = 1) = \prod_{j=1}^m \frac{j}{j + m^\alpha}. \quad (5.38)$$

Na forma mais geral $\theta(s)$ é dado por (veja Figura 5.4)

$$\theta(s) = \prod_{j=1}^m \frac{1}{1 + \frac{s^{1-\alpha} m^\alpha}{j}}. \quad (5.39)$$

Podemos então, escrever a probabilidade B a partir da Equação 5.39 simplesmente multiplicando o $\theta(s)$ de cada instante de tempo s

$$B = \prod_{s=1}^{\infty} \theta(s). \quad (5.40)$$

Aplicando \ln em ambos os lados da equação 5.40 obtemos

$$\ln B = \sum_{s=1}^{\infty} \ln \theta(s), \quad (5.41)$$

Para $s \gg 1$ temos que

$$\ln \left(1 + \frac{m^\alpha}{j s^{\alpha-1}} \right) \sim \frac{m^\alpha}{j s^{\alpha-1}}. \quad (5.42)$$

E então, podemos reescrever a equação 5.41 como

$$\ln B = - \sum_{s=1}^{\infty} \frac{m^\alpha}{s^{\alpha-1}} \sum_{j=1}^m \frac{1}{j}, \quad (5.43)$$

ou ainda

$$\ln \mathbf{B} = - \sum_{s=1}^{\infty} \frac{m^{\alpha} q}{s^{\alpha-1}}. \quad (5.44)$$

onde

$$q = \sum_{j=1}^m \frac{1}{j}, \quad (5.45)$$

e então, assumindo que s varie continuamente podemos escrever $\ln \mathbf{B}$ da seguinte forma

$$\ln \mathbf{B} = - \int_{s=1}^{\infty} \frac{m^{\alpha} q}{s^{\alpha-1}} ds. \quad (5.46)$$

Integrando a equação 5.46 obtemos que

$$\ln \mathbf{B} = - \left. \frac{m^{\alpha} q}{s^{\alpha-2}(2-\alpha)} \right|_{s=1}^{\infty}. \quad (5.47)$$

Da equação 5.47 temos que $\mathbf{B} = 0$ quando $\alpha \leq 2$ e $\mathbf{B} > 0$ caso contrário, reproduzindo os resultados obtidos para $m = 1$ [97, 99].

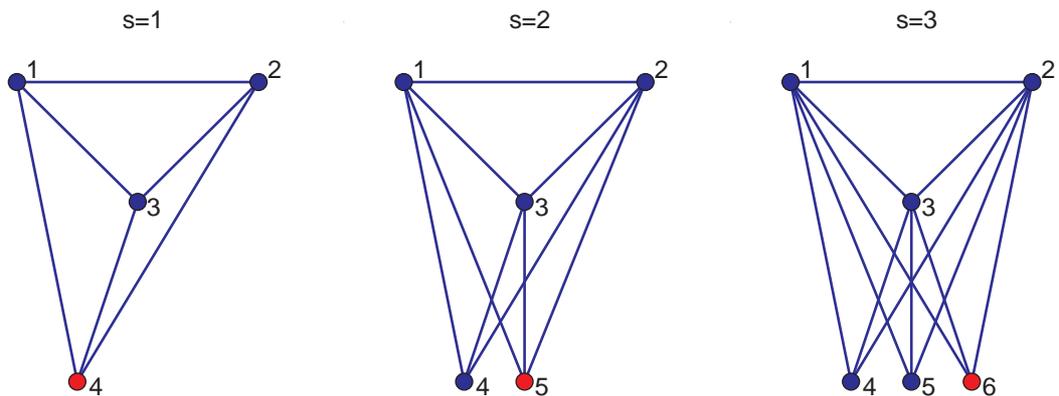


Figura 5.4: Evolução temporal de $\theta(s)$ para os instantes de tempo $s = 1, 2, 3$, respectivamente.

5.2 Distribuição de Conectividades na Aproximação de Campo Médio

No contexto de campo médio, a distribuição de conectividades foi obtida para a rede **BA** linear [14, 65] e **BA** não-linear (com $m=1$) [96]. Nesta seção derivamos a distribuição de conectividades para o modelo **BA** não-linear com m qualquer [99].

Assumindo que a conectividade k_i de um vértice i (de uma rede com N vértices), seja uma variável contínua, é podemos escrever

$$\frac{\partial k_i}{\partial t} = m \frac{k_i^\alpha}{\sum_{j=1}^N k_j^\alpha}. \quad (5.48)$$

Como vimos na seção anterior, quando $\alpha \leq 1$ temos que

$$\sum_{j=1}^N k_j^\alpha = \mu N \sim \mu t, \quad (5.49)$$

e então, substituindo 5.49 em 5.48 e integrando, temos

$$k_i = m \left[1 + \frac{m^\alpha(1-\alpha)}{\mu} \ln\left(\frac{t}{t_i}\right) \right]^{\frac{1}{1-\alpha}}, \quad (5.50)$$

onde t_i é o instante de tempo em que o vértice i é incorporado pela rede e $k_i(t_i) = m$ é a conectividade do vértice i no instante t_i . Assumindo que a probabilidade de que um vértice aleatoriamente escolhido é

$$P(t_i) = \frac{1}{N}, \quad (5.51)$$

então, a Equação 5.50 pode ser vista como uma mudança de variáveis estocástica $t_i \rightarrow k_i$, da forma

$$P(k) = \int \delta(k - k_i(t_i)) P(t_i) dt_i. \quad (5.52)$$

Assim, a distribuição de conectividades é dada por

$$P(k) = \frac{\mu}{m k^\alpha} \exp \left\{ -\frac{\mu}{m^\alpha (1-\alpha)} \left[\left(\frac{k}{m} \right)^{1-\alpha} - 1 \right] \right\}, \quad (5.53)$$

que é o comportamento assintótico do $P(k)$. Na Seção 5.3 usaremos esse resultado de campo médio no cálculo do coeficiente de agrupamento.

Se quisermos construir uma rede de escala livre com expoente arbitrário (diferente de 3), basta fazer uma pequena e simples modificação no modelo BA linear, na qual um vértice com m ligações se conecta um vértice de conectividade k_i da rede proporcional a $k_i + \lambda$. Seguindo esse procedimento utilizado obtemos a seguinte regra de ligação preferencial para o modelo **BA** modificado

$$\frac{\partial k_i}{\partial t} = m \frac{(k_i + \lambda)}{\sum_{j=1}^N (k_j + \lambda)}, \quad \lambda > -m, \quad (5.54)$$

então, a distribuição de conectividades do modelo **BA** modificado será

$$P(k) = \left(2 + \frac{\lambda}{m} \right) \left(\frac{k + \lambda}{m + \lambda} \right)^{-3 - \frac{\lambda}{m}}, \quad (5.55)$$

que, no limite assintótico $k \gg 1$, é regida pela lei de potência $k^{-3 - \frac{\lambda}{m}}$ com o expoente sendo ajustado pelo parâmetro λ .

5.3 Coeficiente de Agrupamento do Modelo BA não-linear

Como vimos no Capítulo 4, o coeficiente de agrupamento é uma propriedade transitiva da rede. A forma mais geral dessa propriedade é o coeficiente de agrupamento de ordem x do vértice i , $C_i(x)$, que é definido como a probabilidade de que dois primeiros vizinhos de um vértice estejam a uma distância de comprimento x (medida sem passar pelo vértice i) [100]. A forma do coeficiente de agrupamento $C_i(x)$ é

$$C_i(x) = \frac{2y_i}{k_i(k_i - 1)}, \quad (5.56)$$

onde y_i é o número de primeiros vizinhos do vértice i que se encontram a uma distância x e k_i é a conectividade do vértice i . Considerando todos os vértices da rede, o coeficiente de agrupamento médio é dado por

$$C(x) = \frac{\sum_{i=1}^N C_i(x)}{N}, \quad (5.57)$$

A definição 5.56 pode ser melhor compreendida com o exemplo da Figura 5.5, onde tomamos como referência o vértice i com $k_i = 5$, que tem como primeiros vizinhos os vértices $\{1, 2, 3, 4, 5\}$. Na Matriz 1 vemos a menor distância entre os vértices adjacentes ao vértice i . As distâncias obtidas da Figura 5.5 (rede à direita).

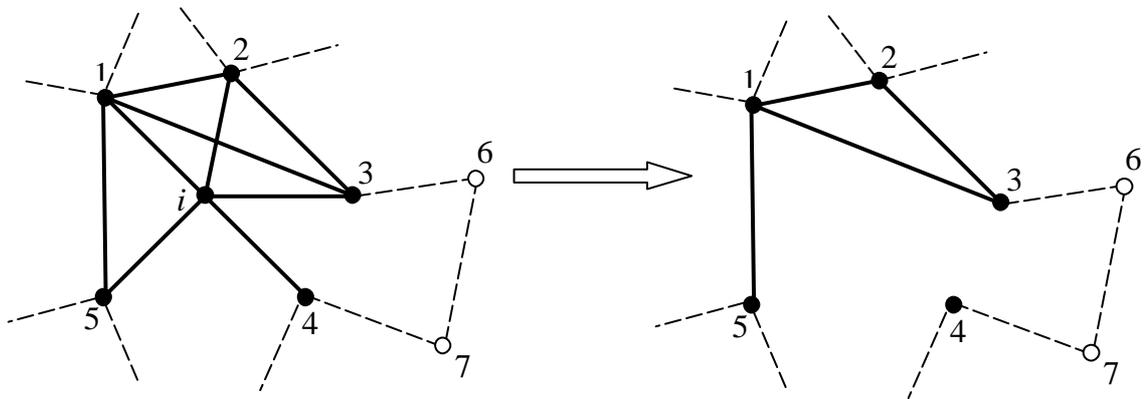


Figura 5.5: Rede na vizinhança do vértice i (rede à esquerda). Rede depois de removido o vértice i e suas ligações (rede à direita).

$$\begin{array}{c}
 \begin{array}{ccccc}
 & 1 & 2 & 3 & 4 & 5 \\
 \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} & \left(\begin{array}{ccccc}
 - & 1 & 1 & 4 & 1 \\
 1 & - & 1 & 4 & 2 \\
 1 & 1 & - & 3 & 2 \\
 4 & 4 & 3 & - & 5 \\
 1 & 2 & 2 & 5 & -
 \end{array} \right)
 \end{array}
 \end{array}$$

Matriz 1: menor distância entre os cinco primeiros vizinhos do vértice i .

Da Matriz 1, podemos calcular o número de vértices a uma distância x e ainda o coeficiente de agrupamento a uma distância x , ambos calculados com relação ao vértice i , e apresentados na Matriz 2,

$$\begin{array}{c}
 x \\
 y_i(x) \\
 C_i(x)
 \end{array}
 \begin{pmatrix}
 1 & 2 & 3 & 4 & 5 \\
 4 & 2 & 1 & 2 & 1 \\
 0.4 & 0.2 & 0.1 & 0.2 & 0.1
 \end{pmatrix}.$$

Matriz 2: $y_i(x)$ – Número de vértices a uma distância x e

$C_i(x)$ – coeficiente de agrupamento a uma distância x .

Aqui iremos tratar apenas o caso $x = 1$, e então podemos abandonar o índice x da Equação 5.56 e usar simplesmente C como sendo o coeficiente de agrupamento médio. Naturalmente, em se tratando da rede **BA** não-linear com $m = 1$, o coeficiente de agrupamento C é sempre igual a zero (independentemente do valor de α) uma vez que a rede é um grafo do tipo árvore.

Podemos obter a expressão analítica para C da rede **BA** não-linear na aproximação de campo médio. Sendo $P(j \rightarrow i)$ a probabilidade de que no tempo $t = j$ o vértice j esteja conectado ao vértice i . Isto significa que

$$P(j \rightarrow i) = m \left[\frac{k_i^\alpha(j)}{\sum_{l=1}^N k_l^\alpha(j)} \right], \quad (5.58)$$

onde $k_i(j)$ é a conectividade do vértice i no instante de tempo j . Para $\alpha \leq 1$, o denominador dessa probabilidade é igual a μj e usando a Equação 5.50 obtemos

$$P(j \rightarrow i) = \frac{m^{1+\alpha} \left[1 + \frac{m^\alpha(1-\alpha)}{\mu} \ln\left(\frac{j}{i}\right) \right]^{\frac{\alpha}{1-\alpha}}}{\mu j} \quad j > i. \quad (5.59)$$

para α igual a 1 (um), a Equação 5.59 é reduzida a

$$P(j \rightarrow i) = \frac{m}{2} (ij)^{-1/2}. \quad (5.60)$$

como já havia sido verificado por Klemm e Eguíluz [101].

Para calcular o coeficiente de agrupamento C , devemos tomar muito cuidado com a ordem dos índices $j > i$, que não podem ser trocados. Se i e j são primeiros vizinhos do vértice l e assumimos que $i < j$, então a contribuição vem dos três regimes: $l < i < j$, $i < l < j$ e $i < j < l$. Assim sendo, o coeficiente de agrupamento pode ser escrito como,

$$C = \frac{1}{N} \left[\sum_{j=3}^N \sum_{i=2}^{j-1} \sum_{l=1}^{i-1} \frac{P(i \rightarrow l)P(j \rightarrow l)P(j \rightarrow i)}{n_l(N)} + \sum_{j=3}^N \sum_{l=2}^{j-1} \sum_{i=1}^{l-1} \frac{P(l \rightarrow i)P(j \rightarrow l)P(j \rightarrow i)}{n_l(N)} + \sum_{l=3}^N \sum_{j=2}^{l-1} \sum_{i=1}^{j-1} \frac{P(l \rightarrow i)P(l \rightarrow j)P(j \rightarrow i)}{n_l(N)} \right], \quad (5.61)$$

onde,

$$n_l(N) = \frac{k_l(N)[k_l(N) - 1]}{2}, \quad (5.62)$$

é o número total de pares de vizinhos que o vértice l tem no instante de tempo (ou tamanho) N . A expressão 5.61 será verificada numericamente logo adiante.

Para $\alpha \leq 1$, o coeficiente de agrupamento vai a zero com N . Existem dois casos particulares nos quais o comportamento assintótico de C é conhecido exatamente: quando $\alpha = 0$ (grafo aleatório) $C \propto \frac{1}{N}$ e quando $\alpha = 1$, $C(N) \propto N^{-1}(\ln N)^2$ [101]. A Equação 5.61 é mais consistente com os resultados da simulação numérica do que com as previsões teóricas anteriores ($C(N) \propto N^{-0,75}$ [102]).

Simulamos o modelo **BA** não-linear em redes de tamanhos $N = 102\,400$ e $m = 2$. Para cada rede, o coeficiente de agrupamento (médio) foi determinado por meio de 20 realizações independentes. Na Figura 5.6, vemos a dependência de C com α para redes de tamanho $N = 102\,400$ e $N = 25\,600$.

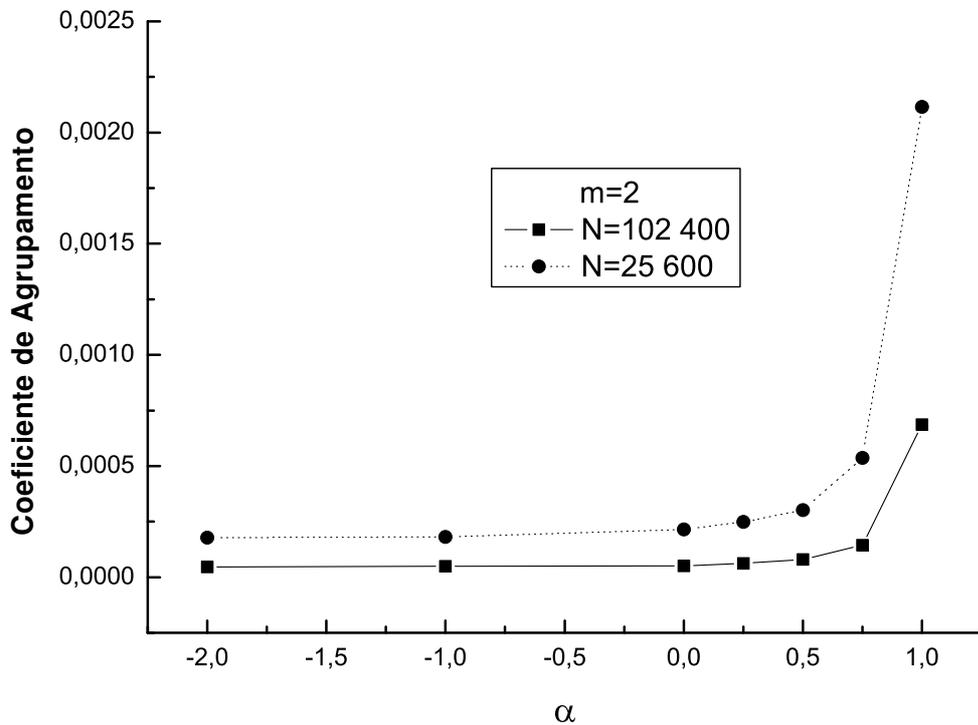


Figura 5.6: O coeficiente de agrupamento médio C vs o expoente α para dois diferentes tamanhos de rede.

Na Figura 5.7, comparamos o resultado da expressão analítica 5.61 com os resultados da simulação do modelo **BA** não-linear.

Para $\alpha > 1$, o valor do coeficiente de agrupamento se aproxima rapidamente de 1 (um) quando o tamanho da rede N cresce. Este resultado vem do fato de que os m vértices super-conectados também são interconectados. Então, quando um novo vértice é adicionado à rede, a forte dinâmica de ligação preferencial ($\alpha > 1$) faz com que ele se conecte aos vértices super-conectados, com isso o valor do coeficiente de agrupamento se aproxima de 1 (um).

5.4 Coeficiente de Assortatividade

Um importante aspecto das redes é a forma como os vértices estão ligados. Se existe uma tendência para conectar vértices com aproximadamente o mesmo

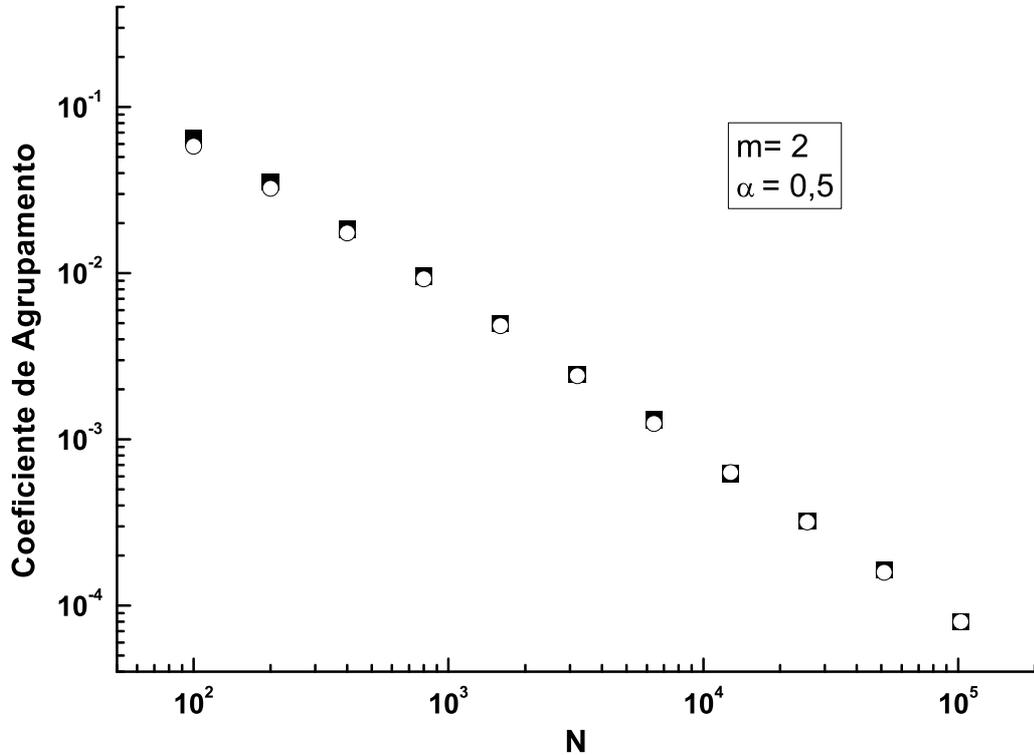


Figura 5.7: Comparação entre simulação numérica e a teoria. Quadrados pretos são os pontos de simulação e os círculos abertos correspondem a formula analítica proposta.

número de ligações, chamamos isso de coeficiente de assortatividade. Quando as ligações favorecem vértices de diferentes conectividades o sistema é dito desassortativo. A situação neutra é chamada de não assortativo [103, 104]. Para medir o coeficiente de assortatividade, fazemos uma pequena mudança na definição dada por Callaway *et al.* [105]. Seja e_{jk} a probabilidade conjunta de que uma ligação aleatoriamente escolhida conecte vértices com conectividade j e k . A distribuição e_{jk} obedece a seguinte regra de soma

$$\sum_{jk} e_{jk} = 1, \quad \sum_j e_{jk} = q_k. \quad (5.63)$$

Podemos definir o coeficiente de assortatividade r [103, 104] como sendo

$$r = \frac{1}{\sigma_q^2} \sum_{jk} jk(e_{jk} - q_j q_k), \quad (5.64)$$

onde, $\sigma_q^2 = \sum_k k^2 q_k - \left[\sum_k k q_k \right]^2$ é a variância da distribuição q_k (veja Capítulo 4. A quantidade r é definida para o intervalo $-1 \leq r \leq 1$, e assume o valor $+1$ (-1) se o sistema é perfeitamente assortativo (desassortativo) e é zero no caso de não assortativo.

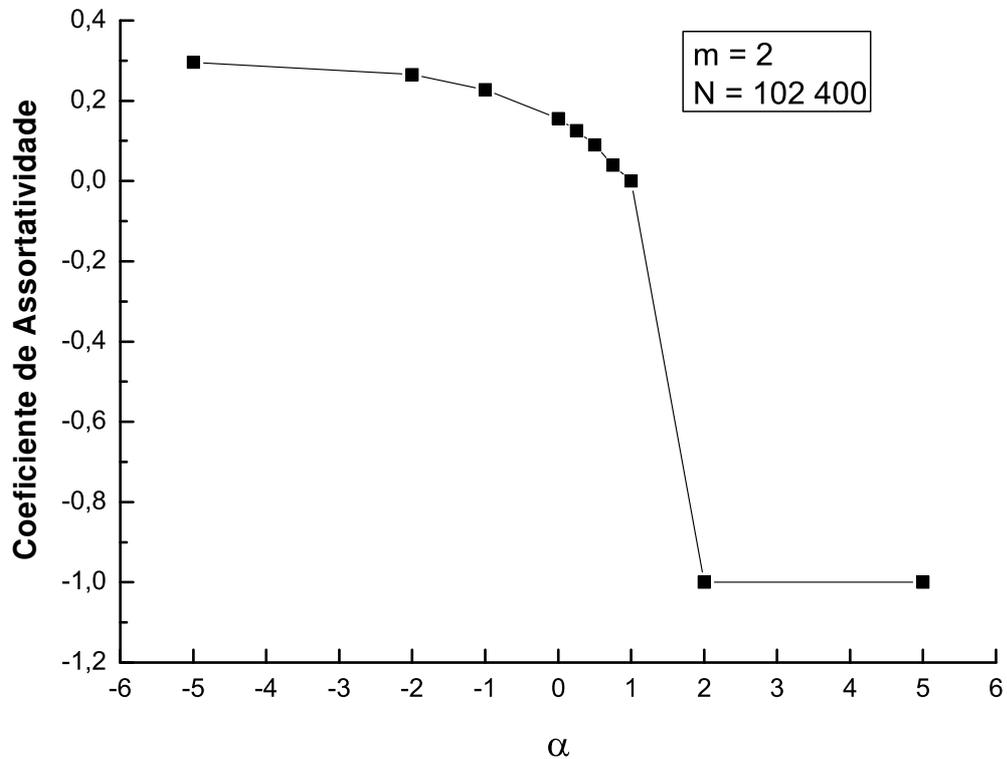


Figura 5.8: A dependência do coeficiente de assortatividade r com o expoente α .

Em nossas simulações (veja Figura 5.8), construímos redes de 102 400 vértices de acordo com o modelo **BA** não-linear, para vários valores de α , e determinamos o valor do coeficiente de assortatividade r de cada rede. Na Figura 5.8 vemos o comportamento do coeficiente de assortatividade r contra o expoente α . No limite de rede infinita,

$r = 0$ se $\alpha = 1$, um resultado já conhecido [103]. Nossos resultados numéricos mostram que esse é o único caso em que $r = 0$. O modelo **BA** não-linear é assortativo ($r > 0$) para $\alpha < 1$ e desassortativo ($r < 0$) quando $\alpha > 1$.

No limite $\alpha \rightarrow -\infty$ o coeficiente de assortatividade pode ser calculado analiticamente. Por exemplo, se $m = 2$ então o número de vértice $N(k)$ com conectividade k é $N(2) = 1$, $N(3) = 6$, $N(4) = N - 7$ e zero para $k \geq 5$. A conectividade média $\langle k \rangle$ é igual a $4 - \frac{8}{N}$. A distribuição de probabilidade conjunta e_{ij} ($i, j = 2, 3, 4$) pode ser facilmente estimada

$$e_{ij} = \frac{1}{N \langle k \rangle} \begin{pmatrix} 0 & 1 & 1 \\ 1 & 10 & 7 \\ 1 & 7 & 4N - 36 \end{pmatrix},$$

Usando a definição 5.64 e o resultado da distribuição e_{ij} (acima) podemos calcular o valor exato do coeficiente de assortatividade como sendo

$$r = \frac{7}{13}. \quad (5.65)$$

5.5 Comprimento Médio do Menor Caminho

A distância entre dois vértices de uma rede é o tamanho do comprimento do menor caminho entre eles, também conhecido como geodésica. Quando medimos as geodésicas de todos os pares dos vértices, o valor médio dessas distâncias corresponde ao que conhecemos como o comprimento médio do menor caminho ℓ . Esse é um

conceito fundamental no estudo de redes e muito importante no campo das redes de comunicações [106] e computadores onde procedimento de roteamento e busca são tarefas comuns.

Para estimar ℓ fizemos uso do valor de ℓ_{min} que é definido como a distância média de todos os vértices, medido a partir do vértice de maior conectividade na rede (se existir mais de um vértice que possua conectividade máxima então é feito um sorteio e escolhido um único vértice). O comprimento médio do menor caminho ℓ está limitado ao intervalo [84]

$$\ell_{min} \leq \ell \leq 2\ell_{min}. \quad (5.66)$$

Determinamos ℓ_{min} para vários valores de α e m simulando o modelo **BA** não-linear em redes com tamanho variando entre $N = 100$ e $N = 102\,400$. Fizemos entre 20 e 24 000 realizações e tomamos a média. A Figura 5.9 mostra a variação de ℓ_{min} com α para dois tamanhos diferentes de rede. No limite $N \rightarrow \infty$, ℓ_{min} é infinito para $\alpha \leq 1$ e igual a 1 quando $\alpha > 1$.

Podemos agora analisar a dependência de ℓ_{min} com N . Da Inequação 5.66, conjecturamos que ambos ℓ e ℓ_{min} escalam com N da mesma forma. Redes aleatórias, tais como as redes Erdős-Rényi e Strogatz-Watts, tem um comprimento médio do menor caminho escalando com $\ell \sim \ln(N)$. Para redes livres de escala, tendo uma distribuição de conectividades do tipo $P(k) \sim k^{-\lambda}$, já está demonstrado que $\ell \sim \ln(\ln(N))$ ($\ell \sim \ln(N)$) se $2 < \lambda < 3$ ($\lambda > 3$) [84], ou seja, elas se caracterizam como sendo uma rede ultra-small-world (small-world). No valor de fronteira $\lambda = 3$, a dependência esperada para o comprimento médio do menor caminho é [84].

$$\ell \sim \frac{\ln(N)}{\ln(\ln(N))}. \quad (5.67)$$

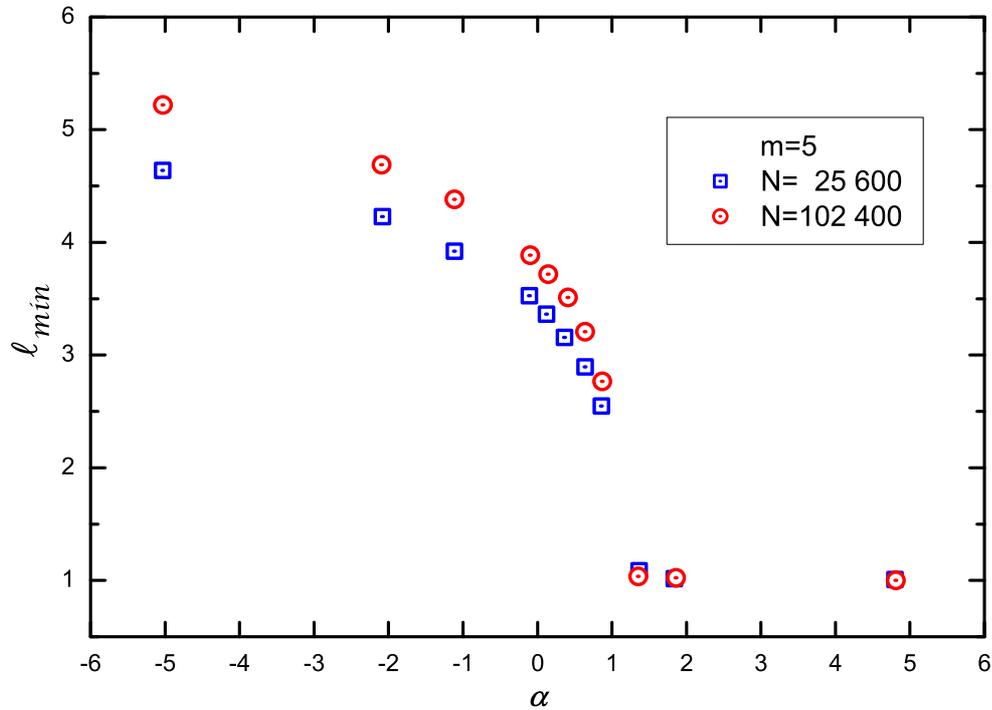


Figura 5.9: O mínimo comprimento médio do menor caminho ℓ_{min} vs o expoente α para dois diferentes tamanhos de rede.

Na aproximação de campo médio, o modelo **BA** linear ($\alpha = 1$) tem um expoente $\lambda = 3$ que independe de m . No modelo **BA** linear, simulado com redes de alguns milhões de vértices, o expoente λ verificado é menor do que 3 [65]. Por meio de uma cuidadosa análise de nossos dados numéricos, obtemos o expoente $\lambda = 2,91 \pm 0,03$. Então, nesse momento, temos a oportunidade de verificar duas previsões teóricas. A primeira é de que se $m = 1$, então o modelo **BA** linear é um grafo do tipo árvore e o comprimento médio do menor caminho escala como $\ell \sim \ln(N)$ [82], e a segunda é a de que se $m \geq 2$, a dependência esperada é $\ell \sim \ln(\ln(N))$ [84]. Como podemos ver nas Figura 5.10, nossos resultados numéricos corroboram as previsões teóricas.

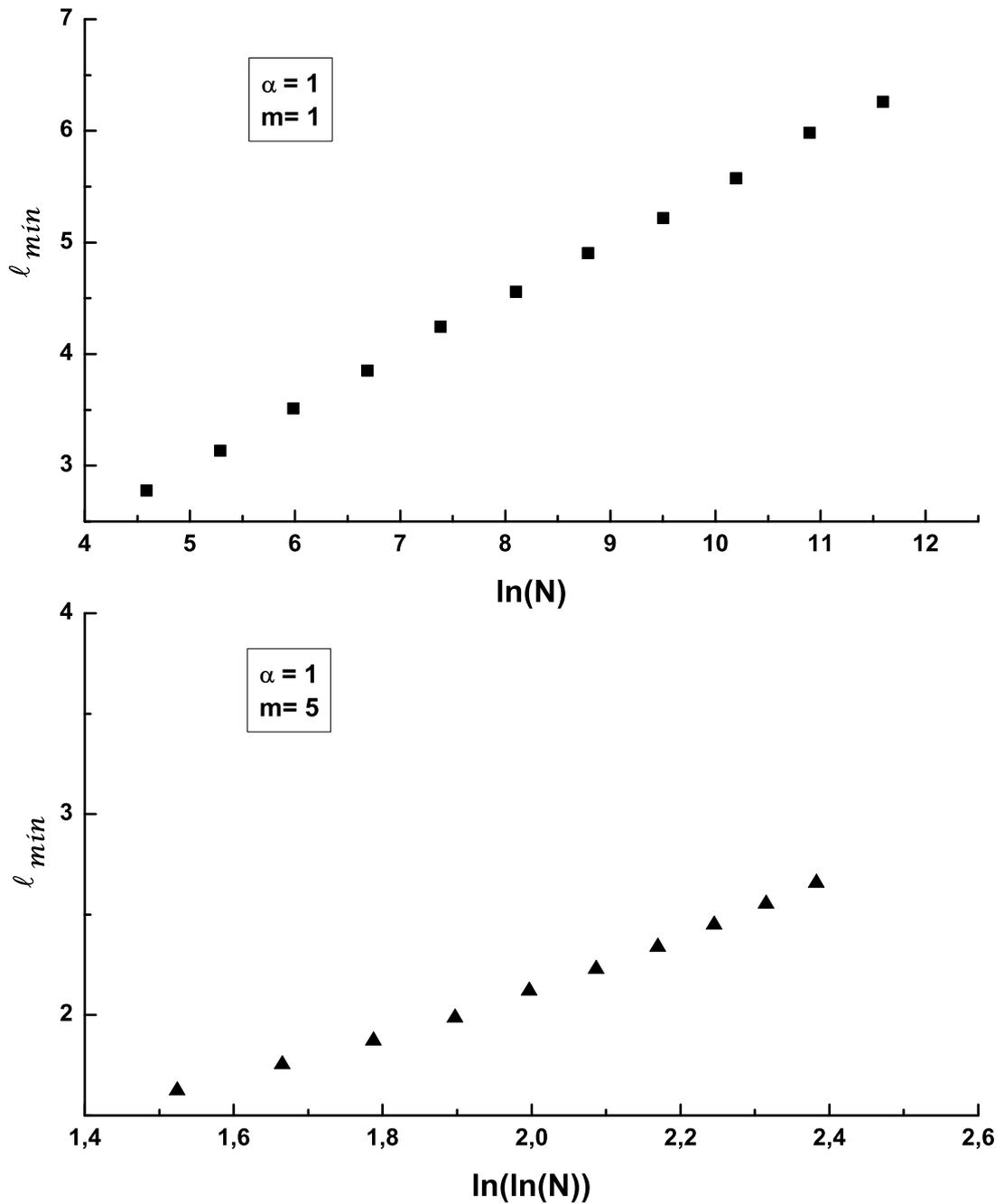


Figura 5.10: O modelo BA linear ($\alpha = 1$). Gráfico mostrando a dependência do mínimo comprimento médio do menor caminho ℓ_{min} vs $\ln(N)$ e $\ln(\ln(N))$ para $m = 1$ e $m = 5$, respectivamente.

5.6 Conclusões

Estudamos algumas propriedades do modelo **BA** não-linear. Usando a equação mestra, obtemos uma expressão analítica para a distribuição de conectividades $P(k)$ para todos os valores de $m \geq 1$ e $\alpha \leq 1$. O setor $\alpha < 0$ tem sido negligenciado, contudo acreditamos que esse setor possa ser útil em estudos de redes sociais, visto que estes sistemas são altamente assortativos, ou podem ser úteis em estudos de uma transição do regime mundo pequeno, onde o comprimento médio do menor caminho cresce logaritmicamente com o tamanho da rede, para o regime mundo grande, onde o comprimento médio do menor caminho cresce rapidamente com alguma potência do tamanho da rede.

No limite $\alpha \rightarrow -\infty$, usando o modelo **BA** não-linear obtivemos uma rede homogênea com conectividade $2m$. Se $\alpha > 1$, existem m vértices super-conectados, caracterizando o que chamamos de fase gel. A probabilidade de que os m vértices iniciais estejam conectados a todos os outros vértices da rede é não nula somente quando $\alpha \geq 2$, sendo que o limiar $\alpha = 2$ não depende de m .

Propusemos uma fórmula analítica para o coeficiente de agrupamento médio C . Sua validade foi verificada por simulações numéricas. Para redes de qualquer tamanho, C é uma função monotonicamente crescente de α . Se $N \rightarrow \infty$, o coeficiente de agrupamento vai a zero para os valores de $\alpha \leq 1$, e se aproxima rapidamente do seu valor máximo $C = 1$ quando $\alpha > 1$.

O padrão associativo do modelo **BA** não-linear foi estudado usando o coeficiente de assortatividade r . Se $\alpha < 1$, então o coeficiente de assortatividade r cresce com o tamanho N da rede e converge assintoticamente para algum valor menor do que 1 (um). Se $\alpha = 1$, r decresce com o tamanho N da rede, chegando a zero no limite de uma rede infinita. Se $\alpha > 1$, r diminui com o tamanho da rede, mas agora converge para o valor -1 . Fazendo um breve resumo, a rede **BA** não-linear é assortativa (desassortativa) se

$\alpha < 1$ ($\alpha > 1$) e não assortativa somente quando $\alpha = 1$, sendo que no limite $\alpha \rightarrow -\infty$ o coeficiente de assortatividade pode ser calculado exatamente, o valor encontrado foi $r = 7/13$ quando $m = 2$.

Fazendo uso de ℓ_{min} estimamos o diâmetro do modelo **BA** não-linear. Para rede de tamanhos fixos, ℓ_{min} é uma função de α monotonicamente decrescente. O valor de ℓ_{min} tende ao infinito se $\alpha \leq 1$ e é igual a 1 quando $\alpha > 1$. No caso particular $\alpha = 1$, caso linear, $\ell_{min} \sim \ln(N)$ se $m = 1$ e $\ell_{min} \sim \ln(\ln(N))$ quando $m \geq 2$.

CAPÍTULO 6

A Rede Complexa do Futebol Brasileiro

6.1 Física no Futebol

*F*utebol, um dos esportes mais populares do mundo, nas últimas décadas tem atraído a atenção da comunidade científica. Num recente artigo, Aguiar e Rubini [107] estudaram duas famosas jogadas, a primeira é a jogada do “gol que Pelé não fez”¹, o gol que Pelé perdeu na Copa de 1970, no jogo contra a então Tchecoslováquia. Na segunda jogada, a “folha seca”² de Valdir Pereira (Didi, o Príncipe Etíope), ainda hoje se discute quais efeitos de fato estão presentes nessa jogada. Neste capítulo, apresentaremos um estudo do futebol do ponto de vista de redes complexas. Dados *reais* foram obtidas de 32 edições do campeonato de futebol brasileiro, totalizando 13 411 jogadores de futebol e 127 clubes. Primeiro consideramos uma rede com dois tipos de vértices, denominada rede bipartida: os jogadores de futebol e os clubes. Num segundo momento consideramos uma rede com um único tipo de vértice, os jogadores. No decorrer desse capítulo apresentaremos alguns resultados interessantes e outros inesperados obtidos no estudo dessas redes. Exemplos desses resultados são as distribuições de probabilidades de que um jogador do futebol

¹Em que se fez presente o efeito Magnus, que surge quando a bola gira em torno de seu eixo perpendicular à velocidade.

²Um chute quase que mortal para os goleiros. O lance foi inventado por Didi em 1956, na partida contra o América. Ele estava com uma contusão que não permitia dar os chutes de longa distância da forma normal. Por isso, ele achou um jeito para a dor desaparecer: acertar um pouco abaixo do meio da bola, o que fazia com que a bola caísse inesperadamente, como uma folha seca.

brasileiro tenha trabalhado em N clubes ou de que tenha jogado M partidas onde ambas apresentam decaimento exponencial. Por outro lado, a distribuição de probabilidade de que um jogador tenha marcado G gols decai segundo uma lei de potência.

6.2 Rede Bipartida do Futebol Brasileiro

Usando dados coletados do CD-ROM da revista Placar [108] construímos uma rede (real) bastante peculiar, a rede do futebol brasileiro. Primeiro construímos uma rede bipartida: o primeiro tipo de vértice representa os clubes, em número de 127, e o segundo tipo de vértice representa os jogadores, que totalizam 13 411. Esses números correspondem ao número total de clubes e jogadores de futebol que participaram pelo menos uma vez do campeonato brasileiro de futebol entre os anos de 1 971 a 2 002³. A rede bipartida (veja Figura 6.1) é construída da seguinte maneira: se um jogador defendeu um clube, então estabelecemos uma ligação entre eles. Procedimentos semelhantes de construção já foram feitos para as seguintes redes bipartidas: ator-filme [31, 109], cientista-artigo [103], diretor-empresa [93] e compositor-interprete [110].

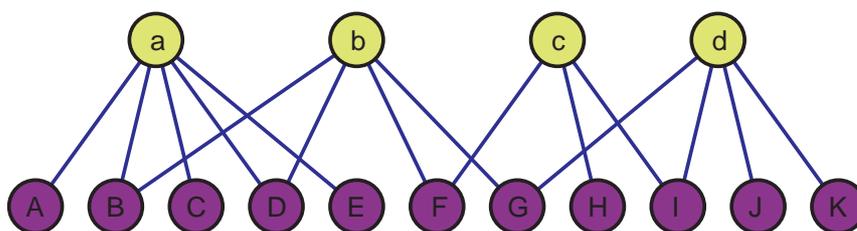


Figura 6.1: Rede bipartida de futebol, com 4 vértices do tipo clube e 11 vértices do tipo jogador de futebol. Vértices de **a** a **d** representam clubes e os vértices de **A** até **K** representam jogadores, e as linhas ligam os jogadores aos clubes que defenderam.

³O primeiro campeonato brasileiro oficial foi disputado em 1 971.

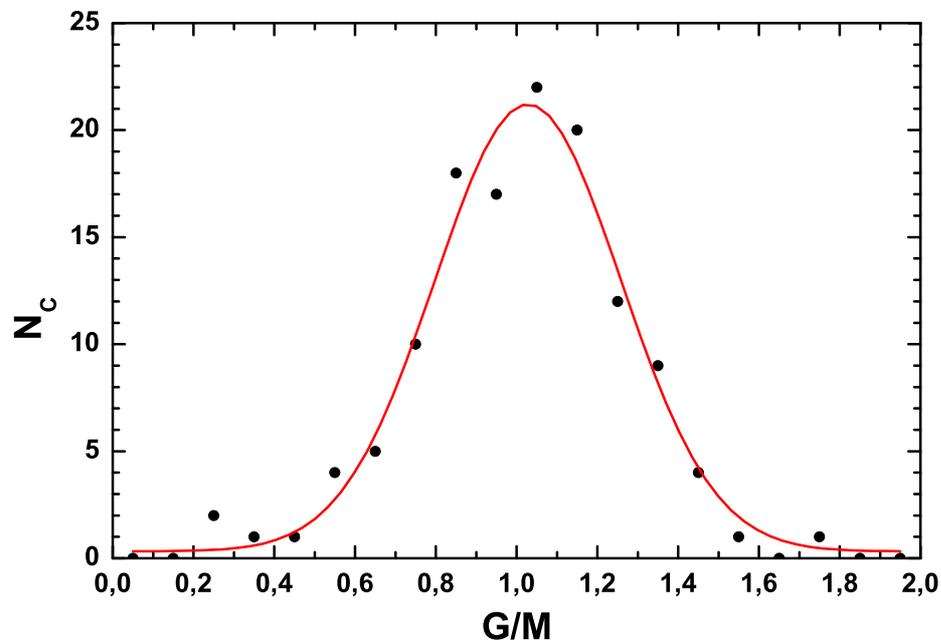


Figura 6.2: Histograma do número de clubes contra o número de gols marcados por partida. A linha cheia corresponde ao ajuste gaussiano. O número médio de gols por partida é aproximadamente 1,0.

Iniciamos nosso estudo analisando a distribuição de gols marcados por partida em função do vários clubes que já participaram do campeonato brasileiro. Na Figura 6.2 vemos o número de clubes N_C versus o número de gols marcados por partida G/M . Como podemos ver os dados são muito bem ajustados por uma gaussiana centrada em $G/M \simeq 1,03$. O clube brasileiro com o melhor desempenho é o São Caetano com uma média de gols por partida de $G/M = 1,73$ e o clube com o menor desempenho é o Colatina com uma média de $G/M = 0,22$.

Analisamos também a distribuição de probabilidades de conectividades para os dois tipos de vértices, jogadores e clubes, da rede bipartida do futebol brasileiro (veja Figura 6.3). Seja N o número de clubes que um jogador já defendeu. Definimos $P(N)$

como a distribuição de probabilidades de que um jogador tenha defendido N clubes diferentes. No gráfico maior da Figura 6.3, vemos que a distribuição de probabilidades de jogadores $P(N)$ exibe um decaimento exponencial com N . Dessa distribuição, determinamos o número médio de clubes que um jogador defendeu, $\bar{N} = 1,37$. Um jogador que merece destaque nessa análise é o Dadá Maravilha, o jogador que mais atuou por diferentes clubes, que participaram do campeonato brasileiro. Ele defendeu 11 diferentes clubes. Seja S o tamanho do clube — que é medido em termos do número de jogadores que atuaram pelo clube — e $P(S)$ a distribuição de probabilidades dos tamanhos dos desses clubes. Vemos no gráfico menor da Figura 6.3, que devido ao pequeno número de clubes, não é possível fazer uma análise conclusiva da distribuição de probabilidades $P(S)$ em função do tamanho S do clube.

Um resultado inesperado é obtido quando determinamos a distribuição de probabilidades $P(M)$ de que um jogador de futebol tenha jogado um total de M jogos pelo campeonato brasileiro (qualquer que tenha sido o clube). Verificamos a existência de um ponto de inflexão (veja Figura 6.4a) ou um limiar nas proximidades de $M_c = 40$. Como existe muita dispersão em $P(M)$, determinamos a sua correspondente distribuição de probabilidades cumulativas $P_c(M)$ [25]. Essa distribuição é bem ajustada por duas diferentes exponenciais: $P_c(M) = 0,150 + 0,857 \times 10^{-0,042M}$ para $M < 40$ e $P_c(M) = 0,410 \times 10^{-0,010M}$ para $M > 40$, como é mostrado na Figura 6.4b. Como sabemos, uma distribuição de probabilidades com formato exponencial gera uma distribuição de probabilidades cumulativa também exponencial e com o(s) mesmo(s) expoente(s) [25]. A existência do limiar M_c indica que depois que um jogador alcançou uma certa fama e notoriedade, provavelmente, será mais fácil para ele se manter como um jogador de futebol, ou seja, um jogador ao cruzar o limiar M_c passou por “um estado probatório” e com isso ganhou estabilidade em seu trabalho.

Seja G o número de gols (a quinta-essência de um jogo de futebol) marcados por um jogador no campeonato brasileiro, e $P(G)$ a distribuição de probabilidades de

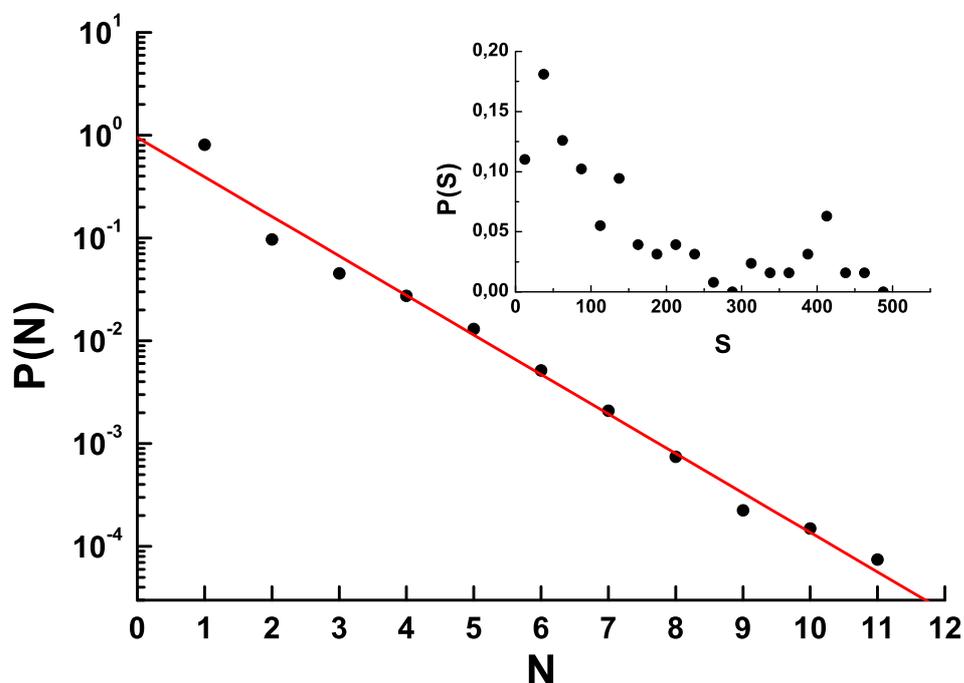


Figura 6.3: Probabilidade $P(N)$ de que um jogador tenha trabalho em N clubes. A linha cheia corresponde ao ajuste da curva $P(N) \sim 10^{-0,38N}$. Assim sendo, vemos que é 190 vezes mais provável de encontrar algum jogador que tenha jogado somente em dois clubes do que encontrar um jogador que tenha jogado em oito clubes. No gráfico menor vemos a distribuição de probabilidade de conectividade $P(S)$ dos clubes.

que um jogador tenha marcado G gols. O comportamento da distribuição de probabilidade $P(G)$ em função do número de gols marcados pode ser visto na Figura 6.5a. Novamente encontramos um limiar intrigante próximo de $G_c = 10$, separando duas regiões que escalam com leis de potências distintas. Um comportamento semelhante já foi observado no contexto de redes de colaborações científicas [25]. Para verificar se esse limiar realmente existe (uma vez que a cauda da curva apresenta muita dispersão) calculamos a distribuição de probabilidade cumulativa correspondente $P_c(G)$ (Figura 6.5b). Assim existe a possibilidade de ajustarmos essa curva com uma lei de potência truncada ou possivelmente duas leis de potências distintas. Nossos melhores resultados foram obtidos fazendo

uso de leis de potências com dois diferentes expoentes: $P_c(G) = -0,259 + 1,256 G^{-0,500}$ para $G < 10$ e $P_c(G) = -0,004 + 4,454 G^{-1,440}$ se $G > 10$.

Lembrando que para uma distribuição de probabilidades escalando com uma lei de potência, a distribuição de probabilidades cumulativa correspondente tem seu expoente acrescido de uma unidade, assim sendo temos que $P(G) \sim G^{-1,5}$ para $G < 10$ e $P(G) \sim G^{-2,44}$ para $G > 10$. Conjecturamos que a origem desse limiar possa ser explicada pelo posicionamento dos jogadores no campo de futebol. Afinal, cerca de dois terços dos onze jogadores formam a defesa ou o meio de campo, conseqüentemente jogadores nessas posições normalmente têm menos chance de fazer gols.

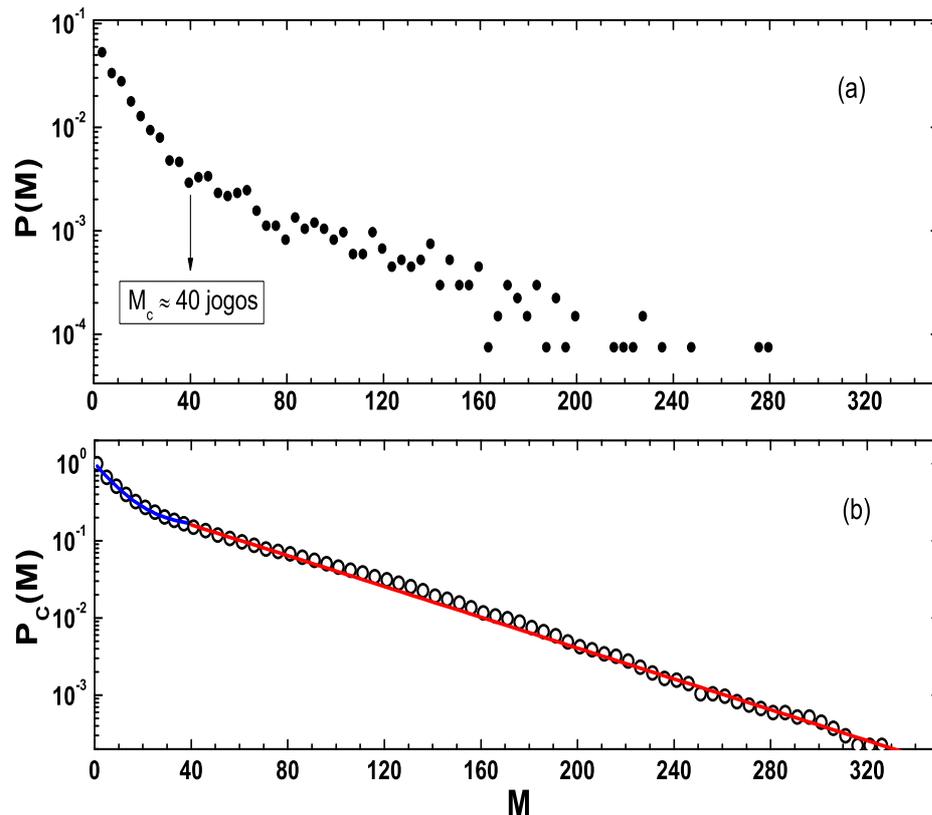


Figura 6.4: (a) Probabilidade de jogos $P(M)$ vs o número M de partidas disputadas. (b) Distribuição de probabilidade cumulativa $P_c(M)$ construída a partir de $P(M)$. As linhas cheias são os ajustes dos conjuntos de pontos.

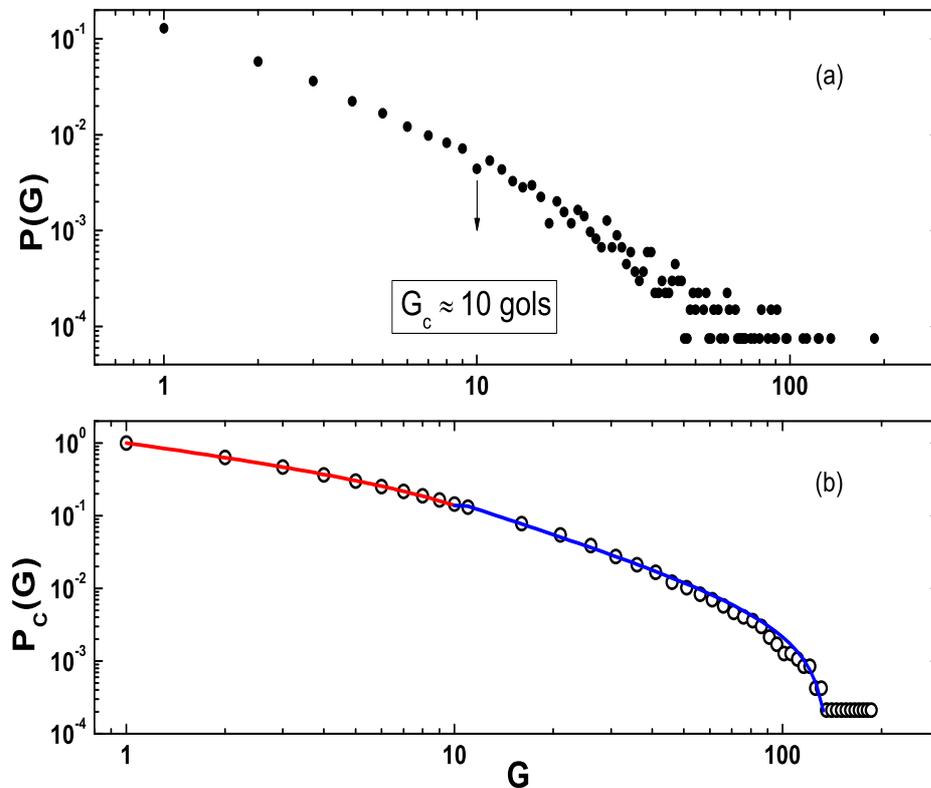


Figura 6.5: (a) Probabilidade $P(G)$ que um jogador tenha marcado G gols. É dez vezes mais provável que um jogador escolhido aleatoriamente tenha marcado 13 gols do que 36 gols. O jogador com o maior número de gols marcados no campeonato brasileiro é o Roberto Dinamite com $G = 186$. (b) Distribuição de probabilidade cumulativa $P_c(G)$.

6.3 Rede Unipartida de Jogadores do Futebol Brasileiro

A partir da rede bipartida do futebol brasileiro (jogadores/clubes), podemos construir a rede unipartida de jogadores (veja Figura 6.6). A construção dessa, que é uma rede real, foi feita da seguinte forma: se dois jogadores defenderam um *mesmo clube* num *mesmo ano* então eles serão unidos por uma ligação. Assim, a rede é composta exclusivamente por vértices que representam jogadores de futebol. Chamaremos esta rede de **JFB** (rede de jogadores do futebol brasileiro). Nessa rede, analisamos a dependência

temporal do comprimento médio do menor caminho, do coeficiente de agrupamento e do coeficiente de assortatividade. Acreditamos que as conexões dos vértices dessa rede representam algum tipo de relação social e/ou profissional estabelecida entre os jogadores.

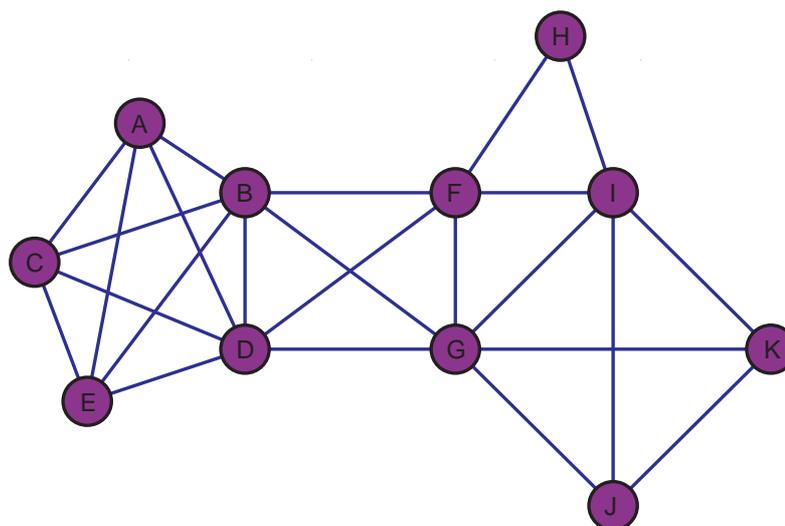


Figura 6.6: Rede unipartida, é a projeção da rede bipartida na qual os jogadores se ligam uns aos outros se eles defenderam um mesmo clube.

No ano de 2002, a rede real **JFB** possuía 13 411 vértices (cada vértice da rede representando um jogador) e 315 566 ligações. A distribuição de probabilidades de conectividades $P(k)$ pode ser facilmente determinada (Figura 6.7), e dela obtemos um valor de conectividade média 47,1 (veja Tabela 6.1), o que significa que um jogador teve 47 colegas de clube, em média.

Outras quantidades da rede **JFB** podem ser exatamente obtidas (veja Tabela 6.1). Por exemplo, a probabilidade C_i de que os primeiros vizinhos do vértice i também estejam conectados. O valor médio dessa quantidade tomado sobre toda a rede é o coeficiente de agrupamento C , um parâmetro muito importante nas redes sociais. Para essa grandeza encontramos o valor de $C = 0,79$, o que significa que a **JFB** é uma rede altamente agrupada.

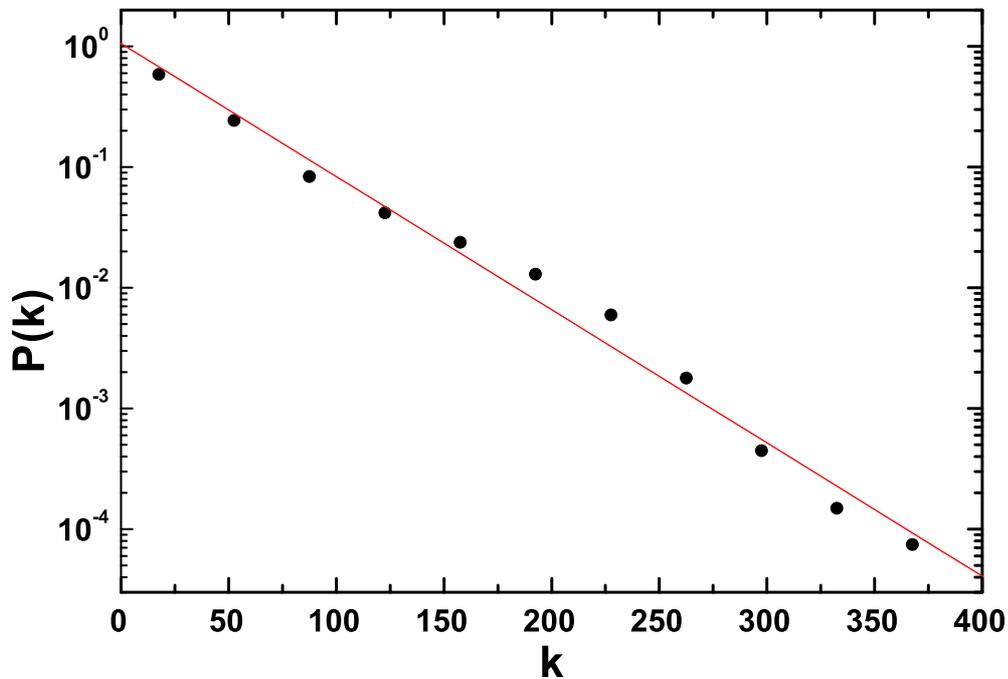


Figura 6.7: Distribuição de conectividades da rede de jogadores do futebol brasileiro, onde um jogador se conecta a outro se eles defenderam um *mesmo clube* num *mesmo ano*. A curva de ajuste (linha cheia) tem a forma exponencial $P(k) \sim e^{-0,025k}$.

Para o comprimento médio do menor caminho $\bar{\ell}$. Quantidade que é obtida medindo as geodésicas de todos os pares de vértices, e depois calculamos o valor médio dessas distâncias. O valor obtido para a rede **JFB** foi $\bar{\ell} = 3,29$. Em analogia com redes sociais, podemos dizer que existe um grau de separação de 3,29 entre os jogadores do futebol brasileiro. Em outras palavras, a rede **JFB** possui característica de uma rede mundo pequeno (veja Capítulo 4).

6.4 Comparação entre as Redes JFB, ER e de Configuração

Como vimos na Seção anterior, a rede **JFB** possui uma distribuição decaindo exponencialmente. Esse é um comportamento inesperado, uma vez que as redes sociais costumam ter a distribuição de conectividade decaindo segundo uma lei de potência. Um bom exemplo é a rede de atores de Hollywood (Figura 6.8). No Capítulo 7 (conclusões) apresentamos uma explicação para a diferença, da forma de decaimento da distribuição de conectividades, entre a **JFB** e as demais redes sociais.

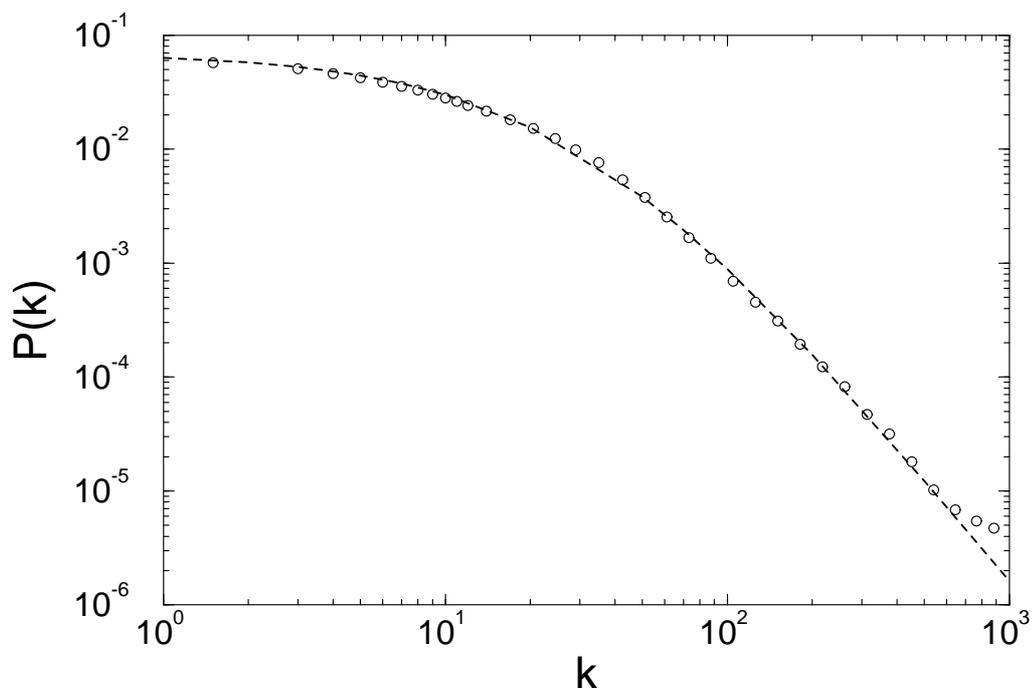


Figura 6.8: Distribuição de conectividades de atores de filmes de Hollywood. Essa figura foi retirado do artigo pela Réka Albert e Albert-László Barabási intitulado *Topology of Evolving Networks: Local Events and Universality* (publicado na revista *Physical Review Letters* **85**, 5234).

A rede **JFB** possui duas características de uma rede aleatória. Distribuição de conectividades decaindo exponencialmente e o valor baixo para o comprimento médio do menor caminho.

Dessa forma, é interessante comparar a rede **JFB** com redes aleatórias que possuam características semelhantes. As redes escolhidas foram o modelo Erdős-Rényi (**ER**) e o modelo de configuração (veja Seções 4.8 e 4.9). Para estabelecer uma maior proximidade entre a rede real **JFB** e as redes **ER** e configuração, simulamos estas duas últimas com o mesmo número de vértices e com um valor de conectividade média aproximadamente igual ao da rede real **JFB**.

No caso do modelo **ER**, construímos a rede usando os valores de 13 411 e 47,1 para o número de vértices e conectividade média, respectivamente. Lembrando da teoria de redes aleatórias, onde temos que $pN = \bar{k}$ (veja Seção 4.8), é fácil ver que o valor da probabilidade de que dois vértices se conectem, usado para construir a rede foi 0,00351.

Para o modelo de configuração, que é uma rede aleatória com uma dada (e fixa) distribuição de conectividades, construímos a rede usando 13 411 vértices e a seguinte distribuição de conectividades $P(k) \sim e^{-0,025k}$, que foi obtida do ajuste da distribuição de conectividades da rede real **JFB** (Figura 6.7).

Na Tabela 6.1, temos as seguintes grandezas: coeficiente de agrupamento, coeficiente de assortatividade e comprimento médio do menor caminho, que foram obtidas para as redes **JFB**, **ER** e de configuração.

Iniciando nossa análise pelo coeficiente de agrupamento, vemos que embora o modelo de configuração tenha um valor de coeficiente de agrupamento duas vezes maior que o do modelo **ER**, esses valores são quase que insignificantes quando comparados com o valor do coeficiente de agrupamento rede **JFB**. Mais uma vez, redes sociais demonstram que serem humanos se agrupam de maneira não aleatória.

Outra quantidade estudada na rede **JFB**, foi o coeficiente de assortatividade A (veja Capítulo 4 e Referência [25]), que mede a tendência de uma rede conectar vértices de mesma conectividade ou de conectividade distinta. Vale lembrar que, se $A > 0$ ($A <$

| | JFB | Erdős-Rényi | Configuração |
|-----------|---------|-------------|--------------|
| v | 13 411 | 13 411 | 13 411 |
| e | 315 566 | 315 443 | 345 294 |
| \bar{k} | 47,1 | 47,1 | 51,5 |
| C | 0,790 | 0,004 | 0,008 |
| A | 0,12 | 0,00 | 0,46 |
| \bar{D} | 3,29 | 2,84 | 2,85 |

Tabela 6.1: Na primeira coluna, v é o número de vértices, e é o número de ligações, \bar{k} é a conectividade média, C é o coeficiente de agrupamento, A é o coeficiente de assortatividade e \bar{D} é o comprimento médio do menor caminho.

0) a rede é dita assortativa (desassortativa) e não assortativa quando $A = 0$. Para a rede **JFB**, encontramos $A = 0,12$, sendo assim concluímos que essa é uma rede assortativa, ou seja, na rede **JFB** jogadores sem grande expressão (de clubes pequenos) no cenário de futebol tendem a jogar com outros jogadores sem expressão e jogadores famosos (de clubes grandes) tendem a jogar outros jogadores famosos.

O valor do coeficiente de assortatividade ($A = 0,12$) da rede JFB, coincide com o resultado obtido para a rede de co-autores em artigos de matemática [103] e é menor do que o obtido da simulação do modelo de configuração ($A = 0,46$). A explicação para isso é a seguinte: apesar dos vértices do modelo de configuração serem conectados aleatoriamente, o vínculo da forma adotada e assumida para a distribuição de probabilidades gera forte correlação. Esta correlação pode ser comprovada pela conectividade média dos primeiros vizinhos de um vértice com conectividade k , $\bar{k}_{nn}(k)$ [87, 111], como apresentado na Figura 6.9.

6.5 Análise da Evolução Temporal da Rede JFB

Além dos estudos da rede bipartida do futebol brasileiro e da rede **JFB**, feitos nas Seções 6.2 e 6.3, analisamos também o crescimento do número de vértices da

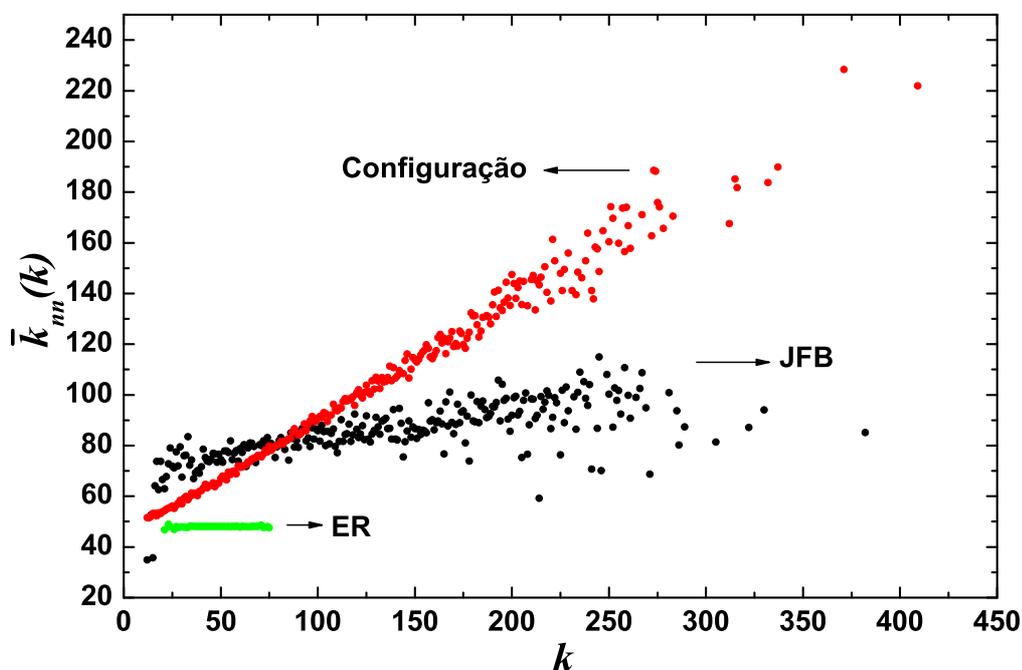


Figura 6.9: Conectividade Média dos Primeiros Vizinhos das redes de **Configuração**, **JFB** e **ER**.

JFB, isto é, a evolução temporal da rede.

Constatamos que no início, 1 971 e 1 972, a rede **JFB** tem várias componentes (veja a Seção 4.5). Porém, a partir de 1 973 a **JFB** possui uma única componente. Na Tabela 6.2 vemos a evolução temporal das quantidades estudadas da rede **JFB**. Para a conectividade média \bar{k} , observamos um crescimento da conectividade média. Podemos pensar em duas razões para isso. A primeira reflete o fato de que a vida profissional do jogador está cada vez mais longa. Dessa forma cada jogador passa a ter mais ligações. A segunda razão é o aumento da frequência com que os jogadores são transferidos entre clubes nacionais.

Os valores do coeficiente de agrupamento, ao contrário dos valores da conectividade média, vem decrescendo com o tempo. Para esse comportamento existe uma

possível razão: êxodo dos (melhores) jogadores de futebol para clubes do exterior, que tem crescido muito nas últimas décadas.

| | 1 975 | 1 980 | 1 985 | 1 990 | 1 995 | 2 002 |
|-----------|--------|---------|---------|---------|---------|---------|
| v | 2 490 | 6 420 | 8 797 | 10 329 | 11 629 | 13 411 |
| e | 48 916 | 128 424 | 181 293 | 219 968 | 254 371 | 315 566 |
| \bar{k} | 39,3 | 40,0 | 41,2 | 42,6 | 43,7 | 47,1 |
| C | 0,84 | 0,83 | 0,82 | 0,81 | 0,80 | 0,79 |
| A | 0,02 | 0,06 | 0,06 | 0,07 | 0,08 | 0,12 |
| D | 3,17 | 3,35 | 3,39 | 3,27 | 3,28 | 3,29 |

Tabela 6.2: Evolução temporal de algumas quantidade da rede **JFB**. O significado das grandezas na primeira coluna são os mesmos da Tabela 6.1.

Na Tabela 6.2, vemos que a rede **JFB** está se tornando mais assortativa com passar dos anos. Isso indica uma possível existência de crescimento de um padrão segregacionista, onde a transferência de jogadores ocorre (veja Figura 6.10), preferencialmente, entre clubes de mesmo tamanho. Finalmente, os valores do comprimento médio do menor caminho sugere uma não dependência com o tamanho da rede. Porém, como foram analisadas poucas gerações de jogadores de futebol, há que se esperar mais tempo para que se verificar de que este padrão é permanente.

CAPÍTULO 7

Conclusões e Perspectivas

A beleza da Física está na simplicidade de suas leis fundamentais. As leis de Newton, as equações de Maxwell, de Schrödinger e de Einstein, todas elas têm algo em comum. Todas podem ser expressas em poucas linhas. No entanto, a natureza se manifesta de maneira muito complexa. Tomemos como exemplo os organismos vivos, que têm as células como unidades básicas. Apesar de termos um bom conhecimento sobre as células, ainda existem perguntas sem respostas a respeito de como o comportamento coletivo das células podem dar origem ao maravilhoso fenômeno da vida. Em outras palavras, o todo parece não se comportar como uma simples superposição de suas partes.

Graças a aguda curiosidade do ser humano, veio à tona a seguinte questão: como o mundo complexo emerge das leis da Física? A busca pela resposta dessa e de outras questões, deram origem ao desenvolvimento da Mecânica Estatística, a partir da segunda metade do século XIX. Em linhas gerais, a Mecânica Estatística é uma área da Física que visa descrever propriedades (complexas) macroscópicas da matéria a partir das interações entre seus constituintes microscópicos. Os constituintes podem ser átomos, moléculas, células, etc.. A complexidade de um sistema se origina das interações não-lineares entre seus constituintes.

Para estudar sistemas complexos, Per Bak e Kim Sneppen, propuseram um modelo simples que exhibe um processo chamado criticalidade auto-organizada — auto-organizada porque não há nenhum parâmetro externo operando sobre o sistema, e crítico

porque as correlações espacial e temporal, do sistema, decaem seguindo uma lei de potência.

Com grande freqüência encontramos fenômenos e sistemas na natureza regidos por uma lei de potência. Exemplos fenômenos naturais regidos por lei de potência são os terremotos, avalanches ou ainda eventos de extinção, como os estudados na década de 1980 pelos paleobiólogos J. J. Sepkoski Jr. e D. M. Raup.

No Capítulo 2, tendo a criticalidade auto organizada como pano de fundo, fizemos simulações de algumas dinâmicas (abaixo citadas) em redes unidimensionais. Estudamos a eficiência das dinâmicas Bak-Sneppen **BS** (e algumas de suas variantes) e a Otimização Extrema Generalizada biológica (**GEO-b**), que é ligeiramente diferente da Otimização Extrema Generalizada **GEO**. A diferença entre estas duas últimas dinâmicas é muito sutil. Enquanto na dinâmica **GEO**, o valor de adaptabilidade λ_i é calculado a partir da hamiltoniana do sistema, na dinâmica **GEO-b**, o valor de adaptabilidade λ_i é obtido de uma distribuição uniforme.

Estudamos a eficiência dos algoritmos **BS** e **GEO-b**, comparando seus valores médios de adaptabilidade no regime estacionário. Com base nos resultados obtidos, concluímos que a dinâmica **BS** é mais eficiente do que a **GEO-b**, no intervalo $[0, 0,86]$ de τ , onde τ é um número real positivo. Para valores de $\tau > 0,86$ a dinâmica **GEO-b** é mais eficiente que a **BS**.

A dinâmica **GEO-b** é acompanhada de algumas características indesejáveis: a correlação espacial entre as espécies é constante, ou seja, ela independente da distância entre espécies subseqüentemente extintas; a existência um parâmetro externo τ prefixado manualmente; a inexistência de equilíbrio entrecortado. A dinâmica **GEO-b** não conduz o sistema a um estado crítico auto-organizado.

Ainda sobre as dinâmicas **BS** e **GEO-b**, podemos dizer que a primeira é uma descrição, simplificada, da evolução biológica seguida pela natureza enquanto a segunda

representa uma otimização criada pelo homem. Ao contrário do que acontece com a dinâmica **BS**, a dinâmica **GEO-b** não tem a característica co-evolucionária, ou seja, a extinção de uma espécie não influencia no destino de espécies vizinhas. A dinâmica **GEO-b** também não possui a característica de eliminação Darwiniana, uma vez que a espécie eliminada não é necessariamente a menos adaptada.

Estudamos também a forma discreta do modelo **BS**, que foi proposto por nós e que chamamos de Bak-Sneppen Discreto (**BSD**). Nesse modelo, discretizamos os valores de adaptabilidade das espécies, afim de verificar se a biodiversidade de espécies é um requisito essencial para alcançar (ou manter) a criticalidade auto-organizada. Podemos interpretar essa discretização como uma diminuição da biodiversidade do sistema.

A diferença entre as dinâmicas **BSD** e **BS** está na atribuição dos valores de adaptabilidades das espécies. Enquanto na dinâmica **BS** o valor de adaptabilidade, λ_i , de cada espécie é sorteado de uma distribuição uniformemente distribuída entre 0 e 1, na dinâmica **BSD** o valor de adaptabilidade λ_i é discreto. Os valores de λ_i são obtidos fazendo $\lambda_i = q/Q$, onde Q (um número inteiro) representa o número de adaptabilidades possíveis e q é um número aleatório que pode assumir os valores 1, 2, ..., ou Q .

Constatamos que, a dinâmica **BSD** pode não possuir a criticalidade auto-organizada. Dessa forma, verificamos que assim como na natureza a biodiversidade parece ter um papel fundamental em modelo evolutivos.

Com o intuito de descobrir o limiar da transição da dinâmica **BSD** para a **BS**, incorporamos um ruído η ao valor da adaptabilidade, produzindo uma nova dinâmica, que chamamos de Bak-Sneppen com Ruído (**BSR**).

Analisamos o modelo **BSR** em busca do valor de ruído a partir do qual a criticalidade auto-organizada emerge no sistema. Como vimos no Capítulo 3, para valores de η menores do que 10^{-7} a dinâmica **BSR** apresenta criticalidade auto-organizada. Interpretamos o ruído η como o surgimento de sub-espécies, que contribuem para o restabelecimento da biodiversidade do sistema.

A partir do Capítulo 4, nos voltamos para o estudo das redes complexas. Apresentamos inicialmente, uma breve revisão histórica do nascimento da teoria de grafos. Revisamos importantes modelos tais como: Erdős-Rényi, Watts-Strogatz, de configuração e o Barabási-Albert, chegando aos dias de hoje com as várias aplicações das redes no estudo de sistemas reais. Algumas propriedades estatísticas dessas redes, tais como comprimento médio do menor caminho, distribuição de conectividades, coeficiente de agrupamento e coeficiente de assortatividade, receberam nossa particular atenção.

O modelo de escala livre **BA** proposto por Albert-László Barabási e sua então aluna de doutorado Réka Albert, contribuiu muito para o desenvolvimento da área de redes complexas. O motivo é que o modelo **BA** captura duas características de redes reais. A primeira é que a rede está em crescimento. A segunda característica é a ligação preferencial. Essa característica em redes sociais é conhecida como *rich get richer* (os ricos tornam-se cada vez mais ricos).

No Capítulo 5, estudamos a forma generalizada do modelo **BA**, o modelo **BA** não-linear. No caso linear ($\alpha = 1$), a rede tem algumas características desejáveis tais como ser de escala livre, e apresentar o efeito mundo pequeno, mas também possui alguns outros aspectos pouco desejáveis tais como, não ser assortativa e ter o coeficiente de agrupamento indo a zero com o aumento da rede.

Além de uma abordagem numérica da dinâmica, fizemos também uma abordagem analítica para o modelo **BA** não-linear, usando a equação mestra. Obtivemos uma expressão analítica para a distribuição de conectividades $P(k)$ para todos os valores de $m \geq 1$ e $\alpha \leq 1$. O setor $\alpha < 0$, que até então havia sido um pouco negligenciado, também foi estudado por nós. Acreditamos que a compreensão desse setor possa ser útil nos estudos de redes sociais, visto que sistemas nessas condições são altamente assortativos.

Propusemos ainda, para o modelo **BA** não-linear, uma fórmula analítica para o coeficiente de agrupamento médio C . Sua validade foi estudada por simulações

numéricas. Verificamos que para redes de qualquer tamanho, C é uma função monotonicamente crescente de α . Quando $N \rightarrow \infty$, o coeficiente de agrupamento vai a zero para todo $\alpha \leq 1$, e se aproxima rapidamente do seu valor máximo $C = 1$ quando $\alpha > 1$.

Outra grandeza que estudamos no modelo **BA** não-linear, e que merece destaque, é o coeficiente de assortatividade r . Vale lembrar que, analisando r podemos descobrir informações sobre o padrão associativo do modelo. Isto é, se existe alguma preferência de vértices de baixa conectividade se conectarem a vertices alta ou baixa conectividade.

Verificamos que para $\alpha < 1$, o coeficiente de assortatividade cresce com o tamanho N da rede e converge assintoticamente para um valor menor do que 1. Para $\alpha = 1$, o coeficiente de assortatividade decresce com o tamanho N da rede, chegando a zero no limite de uma rede infinita. Para $\alpha > 1$, o coeficiente de assortatividade diminui com o tamanho da rede. Em resumo, a rede **BA** não-linear é assortativa se $\alpha < 1$, desassortativa se $\alpha > 1$, e é não assortativa somente quando $\alpha = 1$.

Quando $\alpha < 0$, o modelo **BA** não-linear perde sua característica de ligação preferencial. Podemos interpretar isso como uma característica oposta à conhecida como *rich get richer* (os ricos tornam-se cada vez mais ricos). Uma vez que, para valor de α muito negativo, cada novo vértice adicionado na rede tende a se conectar a vértices de baixa conectividade, algo que podemos interpretar como uma “tendência socialista”. Ou seja, quando α é muito negativo, a rede construída possui o mesmo número de vizinhos, como se fora uma rede regular.

Nesta tese, seguimos a classificação sugerida por Mark E. J. Newman e dividimos as redes complexas em quatro categorias: tecnológicas, biológicas, sociais e de informação. Várias redes reais têm sido estudadas nessa última década, sendo que algumas delas se tornaram muito conhecidas. Na categoria de redes tecnológicas temos as redes da Internet e de distribuição de eletricidade. Na categoria de redes biológicas, temos o exemplo clássico da rede de proteínas de levedura. Dentre as redes de informação,

que também são conhecidas como redes de conhecimento, temos a rede de citações e a rede World Wide Web (WWW). Por último, temos as redes sociais, que também têm sido estudadas segundo a visão de redes complexas. Dentre as redes sociais mais famosas, podemos citar a rede de co-autoria em artigos, a rede de contatos sexuais, a rede de atores de Hollywood bem como a rede de jogadores do futebol brasileiro, que foi um dos nossos objetos de estudo.

No Capítulo 6, a partir de dados coletados do CD-ROM da revista *Placar*, estudamos uma rede real bastante peculiar, a rede do futebol brasileiro. Inicialmente, construímos uma rede bipartida, onde o primeiro tipo de vértice representa os clubes, em número de 127, e o segundo tipo de vértice representa os jogadores, num total de 13 411. Esses números correspondem ao número total de clubes e jogadores de futebol que participaram pelo menos uma vez do campeonato brasileiro de futebol entre os anos de 1 971 a 2 002.

Analisamos a distribuição de probabilidades de conectividades para os dois tipos de vértices, jogadores e clubes, da rede bipartida do futebol brasileiro. Verificamos que a distribuição $P(N)$, de que um jogador tenha defendido N clubes diferentes, decai exponencial com N . Um jogador que merece destaque nessa análise é o Dadá Maravilha. Dadá foi o jogador que mais atuou por diferentes clubes, que participaram do Campeonato Brasileiro. Ele defendeu nada menos do que 11 diferentes clubes.

Obtivemos, ainda da rede bipartida, alguns outros resultados interessantes. Dentre eles, está o resultado da distribuição de probabilidades $P(M)$ de que um jogador de futebol tenha participado de M jogos pelo campeonato brasileiro (qualquer que tenha sido o clube). Verificamos na distribuição $P(M)$, a existência de um ponto de inflexão ou um valor crítico em $M_c = 40$. A interpretação que encontramos para M_c é a seguinte: depois que um jogador alcançou uma certa fama e notoriedade, provavelmente, será mais fácil para ele continuar sendo um jogador de futebol. Em outras palavras, ao cruzar o

limiar M_c , é como se o jogador tivesse passado por “um estado probatório” e ganhasse estabilidade em seu trabalho.

Também verificamos a existência de um outro limiar quando estudamos a distribuição de probabilidades, $P(G)$, de que um jogador tenha marcado G gols. Encontramos um limiar em $G_c = 10$, separando duas regiões que escalam com leis de potências distintas. A origem desse limiar pode ser explicada pelo posicionamento dos jogadores no campo de futebol. Afinal, cerca de dois terços dos onze jogadores formam a defesa ou o meio de campo, conseqüentemente jogadores nessas posições têm menores chances de fazer gols.

A partir da rede bipartida do futebol brasileiro (jogadores/clubes), construímos a rede unipartida de jogadores, que batizamos de **JFB** (rede de jogadores do futebol brasileiro). Nessa rede, analisamos o comprimento médio do menor caminho, do coeficiente de agrupamento e do coeficiente de assortatividade.

Analisamos detalhadamente a rede **JFB**. Uma das primeiras quantidades analisadas foi a distribuição de conectividades $P(k)$. Verificamos que a rede **JFB** possui uma distribuição $P(k)$ decaindo exponencialmente. Esse é um comportamento inesperado para uma rede social, uma vez que elas costumam ter a distribuição $P(k)$ decaindo segundo uma lei de potência. Uma explicação para essa diferença de comportamento pode estar no tempo em que atores e jogadores permanecem em suas respectivas redes. No caso da rede de atores de Hollywood, a vida profissional é bastante longa, tendo muitos casos de atores atuando em filmes até os seus últimos anos de vida (com 50, 60, 70 e alguns até com 80 anos de idade). Desta forma, os atores que se tornam grandes estrelas, e tem uma vida profissional longa, acabam naturalmente se transformando em *hubs* da rede e contribuindo para que a distribuição $P(k)$ da rede de atores tenha um decaimento segundo uma lei de potência. Enquanto isso, na rede **JFB**, os jogadores mais importante (famosos) do futebol brasileiro acabam sendo “congelados” (deixam de fazer novas conexões) quando são transferidos para clubes de futebol no exterior. Acrescente a isso o

fato de que a carreira dos jogadores de futebol normalmente terminam muito cedo (com 30 e 40 anos de idade), por razões de deficiência técnica ou por contusões irreversíveis. Por essas razões, não surgem *hubs* na rede **JFB** e com isso a distribuição de $P(k)$ decai exponencialmente.

Ainda da distribuição de probabilidades de conectividades, obtivemos para a conectividade média o valor de 47,1. A interpretação para essa quantidade é bastante simples, cada jogador teve em média 47 colegas de clube.

Determinamos também o coeficiente de agrupamento C . Para essa grandeza encontramos o valor de $C = 0,79$, o que significa que a **JFB** é uma rede altamente agrupada. Para o comprimento médio do menor caminho $\bar{\ell}$, o valor obtido foi $\bar{\ell} = 3,29$. Em analogia com redes sociais, podemos dizer que existe um grau de separação de 3,29 entre os jogadores do futebol brasileiro. Em outras palavras, a **JFB** é uma rede mundo pequeno.

Estudamos o padrão associativo da rede **JFB**, usando o coeficiente de assortatividade A . Vale lembrar que, uma rede é dita assortativa ($A > 0$) se seus vértices preferencialmente se conectam a vértices de mesma conectividade, ou seja, vértices de pequena conectividade se conectam a outros vértices de pequena conectividade e vértices de grande conectividade se conectam a outros vértices de grande conectividade. Obtivemos para o coeficiente de assortatividade da rede **JFB**, o valor de 0,12. Assim, podemos concluir que esta rede é assortativa. A explicação para isso está no fato de que jogadores sem grande expressão (de clubes pequenos) no cenário de futebol jogarem com outros jogadores sem expressão e jogadores famosos (de clubes grandes) jogarem com outros jogadores famosos.

Além das quantidades estáticas de redes, estudadas acima, a rede **JFB** nos deu uma oportunidade rara, uma vez que foi possível analisar a rede e suas grandezas sob o aspecto da evolução temporal. Analisando a rede, constatamos que nas duas primeiras

edições do campeonato brasileiro, 1 971 e 1 972, à rede era fragmentada, isto é, possuía várias componentes. Entretanto, a partir de 1 973 todas essas componentes se interligaram formando um único aglomerado.

Analisamos a evolução temporal das seguintes quantidades: conectividade média, coeficiente de agrupamento, coeficiente de assortatividade e comprimento médio do menor caminho. Para a conectividade média, observamos um comportamento de crescimento com o tempo. Podemos pensar em duas razões para isso. A primeira reflete o fato de que a vida profissional do jogador está cada vez mais longa. Dessa forma cada jogador passa a ter mais ligações. A segunda razão é o aumento da frequência com que os jogadores são transferidos entre clubes nacionais.

Os valores do coeficiente de agrupamento, ao contrário dos valores da conectividade média, vêm diminuindo com o tempo. Uma possível explicação para esse comportamento é a de que o êxodo dos (melhores) jogadores brasileiros para o exterior tem aumentado substancialmente nas últimas décadas.

Da análise da evolução temporal do coeficiente de assortatividade, podemos concluir que a **JFB** está se tornando mais assortativa com o passar dos anos. Isso parece indicar a emergência de um padrão segregacionista, onde a transferência de jogadores ocorre, preferencialmente, entre clubes de mesmo tamanho.

Finalmente, os valores do comprimento médio do menor caminho sugere uma não dependência com o tamanho da rede, já que ela cresce e o comprimento médio permanece praticamente inalterado. Porém, como foram analisadas poucas gerações de jogadores de futebol, seria necessário esperar mais tempo para verificar se este padrão é permanente.

Estudar as redes complexas, nesses últimos quatro anos, tem sido muito gratificante. Uma das muitas razões para isso, é possibilidade real, e imediata, de podermos estudar sistemas (redes) presentes no nosso cotidiano. Contudo é preciso lembrar

que, ainda há muito para ser feito nessa nova área da física. Quando comparada com outras áreas do conhecimento, o estudo das redes complexas ainda está engatinhando. A compreensão dos padrões e grandezas estatísticas das redes reais, está no início.

Como disse Mark Newman, "entre outras medidas, nós contamos triângulos em redes, verificamos seqüências de conectividades dos vértices, mas ainda não temos nenhuma idéia se estas são as únicas grandezas importantes a se medir (muito provavelmente não são) ou se elas são as mais importantes". Em outras palavras, ainda não temos certeza se estamos, ou não, formulando as perguntas certas.

Precisamos saber mais sobre a topologia das redes complexas, e como ela influencia a evolução de processos que acontecem nestas redes. Enquanto propriedades como a distribuição de conectividades, são facilmente obtidas, outras propriedades tais como correlação, transitividade e estrutura de comunidade, são difíceis e requerem novos algoritmos.

Dando continuidade ao trabalho de doutorado, pretendemos investigar o modelo co-evolutivo de Bak-Sneppen numa rede Barabási-Albert não-linear.

Assim como fizemos com a rede do futebol brasileiro, pretendemos estudar a rede bipartida de novela-atores de novelas brasileiras, sabemos que já foram realizadas algo em torno de 750 novelas durante os 55 anos das novelas brasileiras.

Finalmente, desejamos estudar a estrutura das redes de distribuição elétrica do estado de São Paulo e da Bahia, usando o nosso conhecimento de redes complexas. Pretendemos propor modelos computacionais que ajudem a tornar os sistemas de distribuição robusto tanto a falhas quanto a ataques intencionais, tornando-os eficientes e economicamente viáveis.

APÊNDICE A

Distribuições Lei de Potência, Zipf e Pareto

Quando uma linha monotonicamente decrescente aparece num gráfico log–log, logo começa a discussão e ouvimos alguém dizer que se trata de “uma **lei de potência**”, em seguida um outro afirma “é uma **Zipf**” e ainda um terceiro que contesta os dois primeiros dizendo ter certeza de que se trata de “uma **Pareto**”. E então surge a dúvida, qual dos três está com a razão? A resposta para essa pergunta é: possivelmente todos os três. A seguir tentaremos esclarecer algumas das confusões que giram em torno dessas três leis de distribuição apresentando exemplos e mostrando quando cada uma das distribuições está presente.

Os três termos, acima citados, são usados para descrever fenômeno onde eventos de escalas grandes são raros e eventos de escalas menores são freqüentes. Um exemplo, acontecem poucos terremotos de intensidade muito alta, e acontecem muitos terremotos de intensidades menores. Existem poucas grandes cidades, e existem muitas cidades pequenas. Existem poucas palavras da língua inglesa tais como “the”, “of” e “and” que são muito freqüentes em um texto, enquanto outras são muito pouco freqüentes. A lei de **Zipf** geralmente faz referência a freqüência y da ocorrência de um evento relativo a sua classificação (seu rank) r numa lista. George Kingsley Zipf, um professor de lingüística de Harvard, foi o primeiro a determinar a freqüência da primeira, da segunda, da terceira, ..., da r -ésima palavra mais comum da língua inglesa. A lei de Zipf afirma que

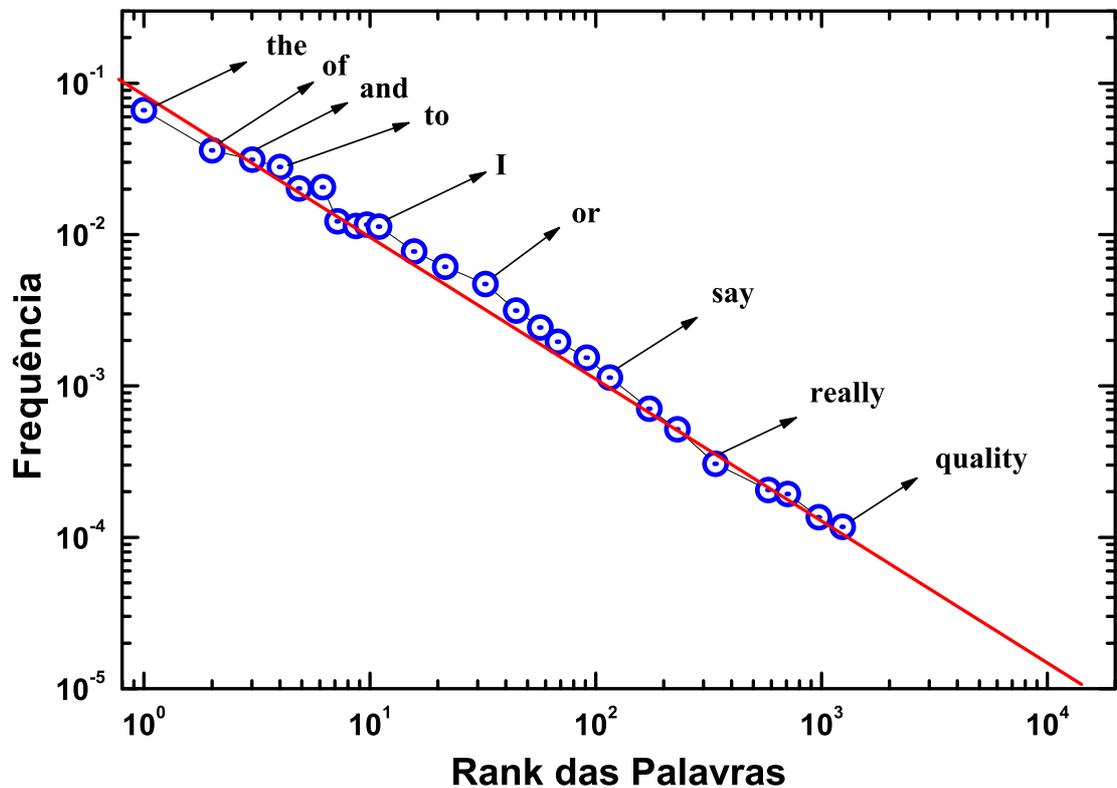


Figura A.1: Ranking de palavras da língua inglesa. A curva mostra que algumas palavras são muito mais frequentes do que outras. Um exemplo é a palavra “the” que é quase 1000 vezes mais frequente do que a palavra “quality”.

a frequência da r -ésima maior ocorrência de um evento é inversamente proporcional ao seu rank:

$$y \sim r^{-b} \quad (\text{A.1})$$

tendo b um valor próximo de 1 (um).

Vilfredo Pareto, um economista e sociólogo italiano, formulou em 1897 sua famosa lei de renda no “Course d’économie politique”. Pareto estava interessado na distribuição de rendas. Ao invés de perguntar qual é a r -ésima maior renda, ele perguntou

quantas pessoas tem uma renda maior ou igual a x . A lei de **Pareto** é dada em termos da distribuição cumulativa, ou seja, o número de eventos maiores do que x é uma potência inversa de x :

$$P[X > x] \sim x^{-k} \quad (\text{A.2})$$

Essa lei afirma, por exemplo, que existem poucos multi-bilionários e que a maioria das pessoas tem uma renda modesta. A lei de Pareto, algumas vezes chamada de distribuição lei de potência, não nos diz quantas pessoas tem uma renda maior do que x , mas sim qual o número de pessoas que tem exatamente uma renda x . Isso é simplesmente a distribuição de probabilidades. É possível mostrar que, existe entre as distribuições cumulativa e de probabilidades uma simples associação, para isso basta fazer

$$P[X = x] x^{-(k+1)} = x^{-a} \quad (\text{A.3})$$

onde $a = 1 + k$ é o expoente da distribuição lei de potência (e k é o parâmetro da distribuição de Pareto).

Apesar da literatura sobre as distribuições de **Zipf** e **Pareto** ser vasta, existem poucos textos que façam uma conexão entre essas distribuições de maneira compreensível, e quanto existe, elas são vagas [1] ou excessivamente complicadas [112, 113].

Para melhor compreender as diferenças entre as distribuições **lei de potência**, **Zipf** e **Pareto**, indicamos dois excelentes artigos. O primeiro artigo é o “Zipfs law and the Internet”, da L. A. Adamic e B. A. Huberman, publicado na revista *Glottometrics* 3, 2002,143-150. O segundo artigo é intitulado “Power laws, Pareto distributions and Zipfs law”, escrito por M. E. J. Newman e publicado na revista *Contemporary Physics* 46, 323-351 (2005).

Bibliografia

- [1] P. Bak. *How Nature Works*. Springer, New York, (1996).
- [2] C. Darwin. *The Origin of Species by Means of Natural Selection*, Murray, London, 1859.
- [3] S. Wright. Character Change, Speciation, and the Higher Taxa. *Evolution*, **36**, 427-443 (1982).
- [4] P. Bak, C. Tang e K. Weisenfeld. Self-Organized Criticality. *Physical Review A*, **38**, 364-374 (1988).
- [5] P. Bak, K. Chen e M. Creutz. Self-organized criticality in the “Game of Life”. *Nature*, **342**, 780 (1989).
- [6] P. Bak, K. Sneppen. Punctuated equilibrium and criticality in a simple model of evolution. *Phys. Rev. Lett.*, **71**, 4083-4086 (1993).
- [7] P. Bak, C. Tang e K. Wiesenfeld. Self-organized criticality: An explanation of the 1/f noise. *Phys. Rev. Lett.*, **59**, 381 (1987).
- [8] S. J. Gould e N. Eldredge. Punctuated equilibria: the tempo and mode of evolution reconsidered. *Paleobiology*, **3**, 115-251 (1977).
- [9] P. Bak e K. Sneppen. Punctuated equilibrium and criticality in a simple model of evolution. *Phys. Rev. Lett.*, **71**, 4083-4086 (1993).
- [10] S. Boettcher e A. G. Percus. Optimization with Extremal Dynamics. *Phys. Rev. Lett.*, **86**, 5211-5214 (2001).
- [11] S. Boettcher e A. G. Percus. Nature’s Way of Optimizing. *Artificial Intelligence*, **119**, 275-286 (2000).
- [12] T. S. Ray e N. Jan. Anomalous approach to the self-organized critical state in a model for “life at the edge of chaos” *Phys. Rev. Lett.*, **72**, 4045-4048 (1994).

- [13] S. Boettcher e M. Paczuski. Exact Results for Spatiotemporal Correlations in a Self-Organized Critical Model of Punctuated Equilibrium. *Phys. Rev. Lett.*, **76**, 348-351 (1996).
- [14] Y. Moreno e A. Vazquez. The Bak-Sneppen model on scale-free networks. *Europhys. Lett.* **57**, 765 (2002).
- [15] L. de Arcangelis e H.J. Herrmann. Self-organized criticality on small world networks. *Physica A*, **308**, 545-549 (2002).
- [16] G. Bianconi e M. Marsili. Clogging and self-organized criticality in complex networks. *Phys. Rev. E*, **70**, 035105(R) (2004).
- [17] T. P. Peixoto e Carmen P. C. Prado. Distribution of epicenters in the Olami-Feder-Christensen model. *Phys. Rev. E*, **69**, 025101(R) (2004).
- [18] K. Park e Y.-C. Lai. Self-organized scale-free networks. *Phys. Rev. E*, **72**, 026131 (2005).
- [19] M. E. J. Newman. The Structure and Function of Complex Networks. *SIAM Review*, **45**, 167-256 (2003).
- [20] Faloutsos, M., P. Faloutsos e C. Faloutsos. On Power-Law Relationships of the Internet Topology. *Computer Communication Review*, **29**, 251-262 (1999).
- [21] Q. Chen, H. Chang, R. Govindan, S. Jamin, S. J. Shenker e W. Willinger. The origin of power laws in Internet topologies revisited. *Proceedings of the 21st Annual Joint Conference of the IEEE Computer and Communications Societies*, IEEE Computer Society (2002).
- [22] R. Govindan, H. Tangmunarunkit. Heuristics for Internet Map Discovery. *Proc. IEEE Infocom 2000*, **3** 1371-1380 (2000).
- [23] H. Jeong, S. Mason, A.-L. Barabási e Z. N. Oltvai. Lethality and centrality in protein networks. *Nature*, **411**, 41-42 (2001).
- [24] S. Redner. How popular is your paper? An empirical study of the citation distribution. *Eur. Phys. J. B*, **4**, 131-134 (1998).
- [25] M. E. J. Newman. The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. USA*, **98**, 404-409 (2001).
- [26] M. E. J. Newman. Scientific collaboration networks: I. Network construction and fundamental results. *Phys. Rev. E*, **64**, 016131 (2001).

- [27] M. E. J. Newman. Scientific collaboration networks: II. Shortest paths, weighted networks, and centrality. *Phys. Rev. E*, **64**, 016132 (2001).
- [28] A.-L. Barabási, H. Jeong, E. Ravasz, Z. Néda, A. Schuberts e T. Vicsek. Evolution of the social network of scientific collaborations. *Physica A*, **311**, 590-614 (2002).
- [29] A.-L. Barabási, H. Jeong, E. Ravasz, Z. Néda, A. Schuberts e T. Vicsek. Evolution of the Social Network for Scientific Collaboration. *Physica A*, **311**, 590 (2002).
- [30] F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley, e Y. Åberg. The web of human sexual contacts. *Nature*, **411**, 907-908 (2001).
- [31] D. J. Watts, e S. H. Strogatz. Collective dynamics of “small-world” networks. *Nature*, **393**, 440-442 (1998).
- [32] L. A. N. Amaral, A. Scala, M. Barthélémy e H. E. Stanley. Classes of small-world networks. *Proc. Natl. Acad. Sci. USA*, **97**, 11149-11152 (2000).
- [33] R. N. Onody e P. A. de Castro. Complex network study of Brazilian soccer players. *Phys. Rev. E*, **70**, 037103 (2004).
- [34] L. da F. Costa, F. A. Rodrigues, G. Travieso e P. R. Villas Boas. Characterization of Complex Networks: A Survey of measurements. Preprint cond-mat/0505185 (2005)
- [35] <http://vlado.fmf.uni-lj.si/pub/networks/pajek/>
- [36] V. Frette, K. Christensen, A. Malte-Sorensen, J. Feder, T. Josang e P. Meakin. Avalanche dynamics in a pile of rice. *Nature*, **379**, 49-52 (1996).
- [37] S. R. Nagel. Instabilities in a sandpile. *Reviews of Modern Physics*, **64**, 321-325 (1992).
- [38] Z. Olami, H. J. S. Feder e K. Christensen. Self-organized criticality in a continuous, nonconservative cellular automaton modeling earthquakes. *Phys. Rev. Lett.*, **68**, 1244-1247 (1992); K. Christensen e Z. Olami. Scaling, phase transitions, and nonuniversality in a self-organized critical cellular-automaton model. *Phys. Rev. A*, **46**, 1829-1838 (1992).
- [39] O. Kinouchi e C. P. C. Prado. Robustness of scale invariance in models with self-organized criticality. *Phys. Rev. E*, **59**, 49644969 (1999).

- [40] J. X. de Carvalho e C. P. C. Prado. Self-Organized Criticality in the Olami-Feder-Christensen Model. *Phys. Rev. Lett.*, **84**, 40064009 (2000).
- [41] R. Burridge e L. Knopoff. Model and theoretical seismicity. *Bull. Seismol. Soc. Am.*, **57**, 341-371 (1967).
- [42] B. Gutenberg e C. F. Richter Frequency of earthquakes in California. *Bull. Seismol. Soc. Am.*, **34**, 185-188 (1944).
- [43] R. N. Onody e P. A. de Castro. Self-Organized Criticality, Optimization and Biodiversity. *International Journal of Modern Physics C*, **14**, 911 (2003).
- [44] H. J. Jensen. Self-Organized Criticality: Emergent Complex Behavior in Physical and Biological Systems. Cambridge University Press, (1998).
- [45] R. K. Sawyer. *Social Emergence: Societies as Complex Systems*. Cambridge University Press, (2005)
- [46] D. M. Raup e J. J. Sepkoski, Jr. Periodicity of Extinctions in the Geologic Past. *Proc. Natl. Acad. Sci. USA*, **81**, 801-805, (1984).
- [47] H. Flyvbjerg, P. Bak e K. Sneppen. Mean field theory for a simple model of evolution. *Phys. Rev. Lett.*, **71**, 4087-4090 (1993).
- [48] Y. M. Pis'Mak. Exact solution of master equations for a simple model of self-organized biological evolution. *J. Phys. A: Math. Gen.*, **28**, 3109-3115 (1995).
- [49] Y. M. Pis'Mak. Solution of the master equation for the Bak-Sneppen model of biological evolution in a finite ecosystem. *Phys. Rev. E.*, **53**, R1326-R1329 (1997).
- [50] J. de Boer, B. Derrida, H. Flyvbjerg, A. D. Jackson e T. Wettig. Simple Model of Self-Organized Biological Evolution. *Phys. Rev. Lett.*, **73**, 906-909 (1994).
- [51] S. Boettcher e M. Paczuski. $d_c = 4$ Is the Upper Critical Dimension for the Bak-Sneppen Model. *Physical Review Letters*, **84**, 2267-2270 (2000).
- [52] S. Kirkpatrick, C. D. Gelatt, e M. P. Vecchi. Optimization by Simulated Annealing. *Science*, **220**, 671-680 (1983).
- [53] J. Holland. *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor (1975).
- [54] P. A. de Castro. *Algoritmos de Otimização e Criticalidade Auto-Organizada*. São Carlos, 2001. 119p. Dissertação (Mestrado) - Instituto de Física de São Carlos, Universidade de São Paulo. (2001).

- [55] L. Pagie e M. Mitchell. A comparison of evolutionary and coevolutionary search. *International Journal of Computational Intelligence and Applications*, **2**, 53-69 (2002).
- [56] A.-L. Barabási. *Linked: The Science of Networks*. Perseus Books Group, (2002).
- [57] S. N. Dorogovtsev e J. F. F. Mendes. *Evolution of Networks*, Oxford University Press (2003).
- [58] P. Erdős e A. Rényi. On random graphs. *Publicationes Mathematicae*, **6**, 290-297 (1959).
- [59] P. Erdős e A. Rényi. On the evolution of random graphs, *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, **5**, 17-61 (1960).
- [60] P. Erdős e A. Rényi. On the strength of connectedness of a random graph. *Acta Mathematica Scientia Hungary*, **12**, 261-267 (1961).
- [61] D. J. Watts. *Small Worlds*. Princeton University Press, Princeton (1999).
- [62] D. J. Watts. Networks, dynamics, and the small world phenomenon. *Am. J. Sociol.*, **105**, 493-592 (1999).
- [63] A.-L. Barabási e R. Albert. Emergence of Scaling in Random Networks. *Science*, **286**, 509-512 (1999).
- [64] A.-L. Barabási, R. Albert e H. Jeong. Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A*, **281**, 69-77 (2000).
- [65] A.-L. Barabási, R. Albert e H. Jeong. Mean-field theory for scale-free random networks. *Physica A*, **272**, 173187 (1999).
- [66] D. J. de S. Price. Networks of scientific papers. *Science*, **149**, 510-515 (1965).
- [67] S. Milgram. The Small-World Problem. *Psychology Today*, **1**, 60 (1967).
- [68] L. A. Adamic e E. Adar. Friends and neighbors on the Web. *Social Networks*, **25**, 211-230 (2003).
- [69] L. A. Adamic e B. A. Huberman. Power Law distribution of the World Wide Web. *Science*, **287**, 2115 (2000).
- [70] Read before you cite!. M. V. Simkin e V. P. Roychowdhury, *Complex Systems*, **14**, 269-274 (2003).
- [71] P. S. Dodds, R. Muhamad e D. J. Watts. An Experimental Study of Search in Global Social Networks. *Science*, **301**, 827-829 (2003).

- [72] P. Sen, S. Dasgupta, A. Chatterjee, P. A Sreeram, G. Mukherjee e S. S. Manna. Small-world properties of the Indian railway network. *Phys. Rev. E*, **67**, 036106 (2003).
- [73] J. R. Banavar, A. Maritan e A. Rinaldo. Size and form in efficient transportation networks. *Nature*, **399**, 130 (1999).
- [74] G. B. West, J. H. Brown e B. J. Enquist. A general model for the structure, and allometry of plant vascular systems. *Nature*, **664** (1999).
- [75] H. D. White, E. Southgate, J. N. Thompson e S. Brenner. The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **314**, 1 (1986).
- [76] S. L. Pimm. *Food Webs*. University of Chicago Press, Chicago (2002).
- [77] J. Camacho, R. Guimerá e L. A. N. Amaral. Robust Patterns in Food Web Structure. *Phys. Rev. Lett.*, **88**, 228102 (2002).
- [78] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai e A.-L. Barabási. The large-scale organization of metabolic networks *Nature*, **407**, 651-654 (2000).
- [79] S. Shen-Orr, R. Milo, S. Mangan e U. Alon. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genetics*, **31**, 64-68 (2002).
- [80] T. I. Lee et al. Transcriptional Regulatory Networks in *Saccharomyces cerevisiae*. *Science*, **298**, 799-804 (2002).
- [81] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai e A.-L. Barabási. Hierarchical organization of modularity in metabolic networks. *Science*, **297**, 1551-1555 (2002).
- [82] B. Bollobas e O. Riordan. *Handbook of Graphs and Networks*. Wiley-VCH, Berlin (2002).
- [83] A. Fronczak, P. Fronczak e J. A. Holyst. Exact Solution for Average Path Length in Random Graphs. Preprint cond-mat/0212230 (2002).
- [84] R. Cohen e S. Havlin. Scale-Free Networks Are Ultrasmall. *Phys. Rev. Lett.*, **90**, 058701 (2003).
- [85] I. J. Farkas, I. Derényi, A.-L. Barabási e T. Vicsek. Spectra of “real-world” graphs: Beyond the semicircle law. *Phys. Rev. E*, **64**, 026704 (2001).
- [86] K. Christensen, R. Donangelo e B. Koiller. Evolution of random networks. *Phys. Rev. Lett.*, **81**, 2380-2383 (1998).

- [87] M. Boguñá, R. Pastor-Satorras e A. Vespignani. Absence of Epidemic Threshold in Scale-Free Networks with Degree Correlations. *Phys. Rev. Lett.*, **90**, 028701 (2003).
- [88] R. Pastor-Satorras, A. Vázquez e A. Vespignani. Dynamical and Correlation Properties of the Internet. *Phys. Rev. Lett.*, **87**, 258701 (2001).
- [89] R. Solomonoff e A. Rapoport. Connectivity of random nets. *Bulletin of Mathematical Biophysics*, **13**, 107-117 (1951).
- [90] E. A. Bender e E. R. Canfield. The asymptotic number of labeled graphs with given degree sequences. *Journal of Combinatorial Theory A*, **24**, 296-307 (1978).
- [91] M. E. J. Newman, S. H. Strogatz e D. J. Watts. Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E*, **64**, 026118 (2001).
- [92] M. Molloy e B. Reed. A critical point for random graphs with a given degree sequence. *Random Structures and Algorithms*, **6**, 161-179 (1995).
- [93] G. F. Davis, M. Yoo, e W. E. Baker. The Small World of the American Corporate Elite, 1982-2001. *Strategic Organization*, **1**, 301326 (2003).
- [94] J. W. Grossman e P. D. F. Ion. On a portion of the well-known collaboration graph. *Congressus Numerantium*, **108**, 129-131 (1995).
- [95] R. de Castro, e J. W. Grossman. Famous trails to Paul Erdős. *Mathematical Intelligencer*, **21**, 51-63 (1999).
- [96] P. L. Krapivsky e S. Redner, Organization of growing random networks. *Phys. Rev. E*, **63**, 066123 (2001).
- [97] P. L. Krapivsky, S. Redner and F. Leyvraz. Connectivity of Growing Random Networks. *Phys. Rev. Lett.*, **85**, 46294632 (2000).
- [98] M. Paczuski, S. Maslov e P. Bak. Avalanche dynamics in evolution, growth, and depinning models. *Phys. Rev. E*, **53**, 414-443 (1996).
- [99] R. N. Onody e P. A. de Castro. Nonlinear Barabási-Albert Network, *Physica A*, **336** (2004) 491502.
- [100] A. Fronczak, J. A. Holyst, M. Jedynek e J. Sienkiewicz. Higher order clustering coefficients in Barabási-Albert networks. *Physica A*, **316**, 688 (2002).
- [101] K. Klemm e V. M. Eguíluz. Growing scale-free networks with small-world behavior. *Phys. Rev. E*, **65**, 057102 (2002).

- [102] R. Albert e A.-L. Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, **74**, 4797 (2002).
- [103] M. E. J. Newman. Assortative Mixing in Networks. *Phys. Rev. Lett.*, **89**, 208701 (2002).
- [104] M. E. J. Newman. Mixing patterns in networks. *Phys. Rev. E*, **67**, 026126 (2003).
- [105] D. S. Callaway, J. E. Hopcroft, J. M. Kleinberg, M. E. J. Newman e S. H. Strogatz. Are randomly grown graphs really random?. *Phys. Rev. E*, **64**, 041902 (2001).
- [106] M. E. J. Newman , S. Forrest e J. Balthrop. Email networks and the spread of computer viruses. *Phys. Rev. E*, **66**, 035101 (2002).
- [107] C. E. Aguiar e G. Rubini. A aerodinâmica da bola de futebol. *Revista Brasileira de Ensino de Física*, **26**, 297 (2004).
- [108] Dados obtidos do CD-ROM da revista Placar. Editora Abril, Brasil (2003).
- [109] L. A. N. Amaral, A. Scala, M. Bartheélémey e H. E. Stanley. Classes of small-world networks. *Proc. Natl. Acad. Sci. USA*, **97**, 11149 (2000).
- [110] D. L. E. Silva, M. M. Soares, M. V. C. Henriques, M. T. S. Alves, S. G. Aguiar, T. P. Carvalho, G. Corso e L. S. Lucena. The Complex Network of the Brazilian Popular Music. *Physica A*, **332**, 559-565 (2004).
- [111] A. Barrat e M. Weigt. On the properties of small-world networks, *Eur. Phys. J. B*, **13**, 547-560 (2000).
- [112] G. Troll e P. beim Graben “Zipf’s law is not a consequence of the central limit theorem”, *Phys. Rev. E*, **57**, 1347-1355 (1998).
- [113] R. Gunther, L. Levitin, B. Shapiro e P. Wagner. “Zipf’s law and the effect of ranking on probability distributions”, *International Journal of Theoretical Physics*, **35**, 395-417 (1996).

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)