



Universidade Federal Rural de Pernambuco  
Pró-Reitoria de Pesquisa e Pós-Graduação  
Departamento de Estatística e Informática

Mestrado em Biometria

## **Evolução Adaptativa em Populações Estruturadas**

Edilson de Araújo Gonçalves

Dissertação de Mestrado  
Recife

junho de 2007

Universidade Federal Rural de Pernambuco  
Pró-Reitoria de Pesquisa e Pós-Graduação  
Departamento de Estatística e Informática

Edilson de Araújo Gonçalves

## **Evolução Adaptativa em Populações Estruturadas**

*Trabalho apresentado ao Programa de Mestrado em Biometria do Departamento de Estatística e Informática da Universidade Federal Rural de Pernambuco como requisito parcial para obtenção do grau de Mestre em Biometria.*

Orientador: *Prof. Dr. Paulo Roberto de Araújo Campos*

Co-orientadora: *Profa. Dra. Viviane Moraes de Oliveira*

Recife

junho de 2007

## Ficha catalográfica

G635e      Gonçalves, Edilson de Araújo  
              Estudo probabilístico da evolução adaptativa em populações estruturadas / Edilson de Araújo Gonçalves. -- 2007.  
              61 f.

              Orientador : Paulo Roberto de Araújo Campos  
              Dissertação (Mestrado em Biometria) - Universidade Federal Rural de Pernambuco. Departamento de Estatística e Informática

              Inclui bibliografia

CDD 574.018 2

1. Evolução adaptativa
  2. Mutações benéficas
  3. Interferência clonal
- I. Campos, Paulo Roberto de Araújo  
II. Título

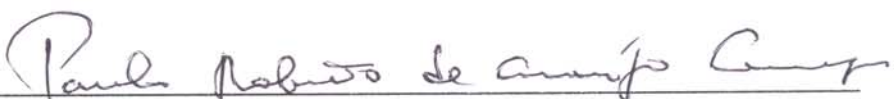
Universidade Federal Rural de Pernambuco  
Departamento Estatística e Informática  
Programa de Pós-Graduação em Biometria

ESTUDO PROBABILÍSTICO DA EVOLUÇÃO ADAPTATIVA  
EM POPULAÇÕES ESTRUTURADAS

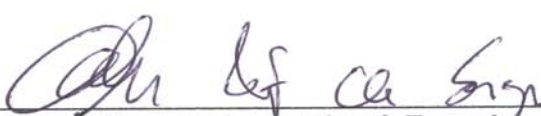
EDILSON DE ARAÚJO GONÇALVES

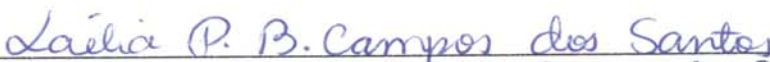
Dissertação julgada adequada para obtenção do título de mestre em Biometria,  
defendida e aprovada por unanimidade em 13/06/2007 pela banca examinadora.

Orientador:

  
Prof. Dr. Paulo Roberto de Araújo Campos  
Universidade Federal Rural de Pernambuco – UFRPE

Banca Examinadora:

  
Prof. Dr. Adauto José Ferreira de Souza  
Universidade Federal Rural de Pernambuco – UFRPE

  
Profa. Dra. Laélia Pumilla Botelho Campos dos Santos  
Universidade Federal Rural de Pernambuco – UFRPE

  
Prof. Dr. Tsang Ing Ren  
Universidade Federal de Pernambuco – UFPE

*Dedico este trabalho aos meus familiares especialmente a meus pais Abdon Vicente de Araújo (in memoriam) e Celina Gonçalves de Araújo que me educaram principalmente na base de exemplos na esperança de que eu viesse a me tornar um cidadão, a meus irmãos, a minha esposa e filhas.*



# Agradecimentos

\* Agradeço a Deus por ter me dado todas as oportunidades.

\* Ao Professor Dr. Paulo Campos pela competência e paciência na condução deste trabalho de orientação.

\* Agradeço ao programa de pós-graduação e a todo seu corpo docente pelo empenho e pela forma profissional como conduziram os trabalhos.

\* A minha família, pai(in memoriam), mãe, irmãos e filhos pelo incentivo recebido.

\* A todos os colegas do curso que certamente contribuíram de uma forma ou de outra para conclusão deste trabalho.

\* Aos Professores do Departamento de Física especialmente ao professor Dr. Ernande Barbosa da Costa pela colaboração e incentivo durante esta caminhada.

\* Ao colega Pedro de Sá Cavalcanti de Albuquerque Neto da iniciação científica pela valiosa contribuição.





*A provação vem, não só para testar o nosso valor, mas para aumentá-lo; o carvalho não é apenas testado, mas enrijecido pelas tempestades.*

—LETTIE COWMAN



# Resumo

A adaptação de populações decorre do surgimento e posterior fixação de mutações benéficas que conferem alguma vantagem seletiva aos indivíduos que as adquirem. Por esta razão, o estudo do processo de fixação de mutações benéficas tem uma longa história na literatura de genética de populações. Particularmente, tais investigações objetivaram encontrar as principais forças evolucionárias que afetam a atuação da seleção natural nas populações. No presente trabalho, nós investigamos a dinâmica de fixação de mutações benéficas em uma população subdividida. As subpopulações, denominadas de demes, podem permutar migrantes entre seus vizinhos. Consideramos duas topologias para a rede de migração, uma tipo grafo aleatório e uma outra para rede livre de escala. Nós verificamos que a taxa de migração afeta drasticamente a dinâmica de fixação de mutações, apesar do fato da probabilidade de fixação ser independente da taxa de migração de acordo com a conjectura de Maruyama. Além disso, observamos uma influência da topologia da rede migratória na evolução adaptativa quando a interferência clonal se torna relevante.

**Palavras-chave:** Evolução adaptativa, mutações benéficas, interferência clonal



# Abstract

Adaptation of populations takes place with the occurrence and subsequent fixation of mutations that confer some selective advantage to the individuals that acquire it. For that reason, the study of the process of advantageous mutations fixation has a long history in the population genetics literature. Particularly, the previous investigations aimed to find out the main evolutionary forces affecting the strength of natural selection in the populations. In the current work, we investigate the dynamics of fixation of beneficial mutations in a subdivided population. The subpopulations (demes) can exchange migrants among their neighbors. The migration network is assumed to have either a random graph or a scale-free topology. We have seen that the migration rate drastically affect the dynamics of mutation fixation, despite of the fact that the probability of fixation is invariant on the migration rate, accordingly to Maruyama's conjecture. In addition, we have observed that a topological dependence of the adaptive evolution of the population exists when clonal interference becomes effective.

**Keywords:** Adaptive evolution, beneficial mutations, clonal interference



# Sumário

<b>1</b>	<b>Introdução</b>	<b>15</b>
<b>2</b>	<b>Redes Complexas</b>	<b>19</b>
2.1	Introdução à Teoria dos Grafos	20
2.2	Modelo de Erdos e Rényi para grafos aleatórios	21
2.3	Rede livre de escala	22
<b>3</b>	<b>Dinâmica Adaptativa</b>	<b>29</b>
3.1	Mutações	30
3.2	Fixação de mutações benéficas	33
3.3	Processo de Ramificação de Galton-Watson	33
<b>4</b>	<b>Modelo de Metapopulações</b>	<b>37</b>
4.1	O modelo Original de Campos e Gordo	40
4.2	Correspondência entre o modelo de metapopulações e o modelo SIS	41
4.3	O modelo modificado com mutações benéficas e deletérias	42
<b>5</b>	<b>Resultados e Discussões</b>	<b>45</b>
5.1	Dinâmica das mutações que não alcançam fixação	51
<b>6</b>	<b>Conclusões</b>	<b>55</b>





# Lista de Figuras

- 2.1 Grafo com  $N = 6$  nós e  $n=5$  arestas. O conjunto de nós é  $P = \{1, 2, 3, 4, 5, 6\}$  e o conjunto de arestas é  $E = \{(1, 3), (3, 5), (3, 4), (5, 6), (4, 6)\}$ . 21
- 2.2 Distribuição de conectividades de um grafo aleatório com  $N = 10000$  nós e (a) conectividade média  $z = 5$  e (b) conectividade média  $z = 10$ . 23
- 2.3 Distribuição de conectividades para uma rede livre de com  $N = 10000$  nós. Rede gerada de acordo com o algoritmo de Albert-Barabási. O expoente obtido da distribuição é  $\gamma = 2.7$ . 27
- 3.1 Probabilidade de fixação  $P_{fix}$  como função da vantagem seletiva  $s_b$ . 36
- 4.1 Ilustração de uma população estruturada. Na figura temos uma população com  $D = 5$  demes, e cada deme possui  $N_d = 10$  indivíduos. 39
- 4.2 Ilustração do processo de fixação de uma mutação benéfica, aqui representada pela cor vermelha. 43
- 5.1 Probabilidade de fixação  $P_{fix}$  como função do valor benéfico  $s_b$  da mutação benéfica. Os dados correspondem a resultados para 5000 simulações independentes. Os valores de parâmetro são  $N = 10000$ ,  $D = 200$ ,  $e = 0$ , e  $m = 0.0001$  (triângulos à esquerda),  $m = 0.001$  (diamantes) e  $m = 0.01$  (círculos). Os símbolos cheios correspondem aos resultados para topologias de grafo aleatório, enquanto que os símbolos vazios se referem a topologias do tipo livre de escala. Em ambos os casos a conectividade média é  $z = 6$ . A linha cheia é predição teórica conforme Eq. (3.11). As barras de erro dos valores estimados via simulação são aproximadamente do mesmo tamanho que os símbolos utilizados. 46

- 5.2 Probabilidade de fixação  $P_{fix}$  como função do valor benéfico  $s_b$  da mutação benéfica. Os parâmetros são os mesmos da Figura 5.1, exceto que  $e = 0.02$ . 47
- 5.3 Probabilidade de fixação  $P_{fix}$  como função do valor benéfico  $s_b$  da mutação benéfica. Os valores de parâmetro são  $N = 10000$ ,  $D = 200$ ,  $e = 0$ , e  $m = 0.0001$  (triângulos à esquerda),  $m = 0.001$  (diamantes) e  $m = 0.01$  (círculos). Os dados são para topologias do tipo grafo aleatório. A conectividade média é  $z = 6$ , e aqui consideramos uma taxa de mutação deletérias não-nula,  $U_d = 0.1$ . 48
- 5.4 Taxa de fixação de mutações benéficas,  $k_b$ , como função da taxa de mutações benéficas  $U_b$ . Os valores de parâmetro são  $N = 10000$ ,  $D = 200$ ,  $e = 0$ ,  $\beta = 20$ , e  $m = 0.0001$  (triângulos á esquerda),  $m = 0.001$  (diamantes) e  $m = 0.01$  (círculos). Os símbolos cheios correspondem aos resultados para topologias de grafo aleatório, enquanto que os símbolos vazios se referem a topologias do tipo livre de escala. Em ambos os casos a conectividade média é  $z = 6$ . A linha cheia corresponde à taxa esperada para uma população homogênea sem interferência clonal conforme Eq.(5.1). 49
- 5.5 Taxa de fixação de mutações benéficas,  $k_b$ , como função da taxa de mutações benéficas  $U_b$ . Os parâmetros são os mesmos da Figura 5.4, exceto que  $e = 0.02$ . 50
- 5.6 Distribuição de probabilidades de tamanhos de população máximos alcançado pelas mutações que foram posteriormente eliminadas. Os parâmetros são  $N = 10000$ ,  $D = 200$ ,  $z = 4$ ,  $m = 0.0001$  e  $e = 0$ . As linhas verticais indicam múltiplos inteiros do tamanho da deme. A topologia da rede é do tipo grafo aleatório. 51
- 5.7 Distribuição de probabilidades de tamanhos de população máximos alcançado pelas mutações que foram posteriormente eliminadas. Os parâmetros são  $N = 10000$ ,  $D = 200$ ,  $z = 4$ ,  $e = 0$  e  $m = 0.0001$  (linha tracejada curta),  $m = 0.001$  (linha tracejada) e  $m = 0.01$  (linha cheia). 52

## Introdução

A adaptação na população de micróbios, no mais simples cenário, consiste em mutações benéficas raras que por meio de rápidas varreduras seletivas alcançam fixação na população. Estes eventos de fixação são alternados por períodos de inatividade, onde nenhuma mutação benéfica é fixada. Embora a ocorrência de mutações deletérias seja muito mais provável, este cenário simples pode ser refutado supondo populações de tamanho muito grande ou elevadas taxas de mutação benéfica. Como consequência dessas circunstâncias, a coexistência de mutações benéficas torna-se comum. Nas populações assexuadas, onde os genomas que carregam mutações benéficas distintas não podem se recombinar para formar uma entidade mais adaptada, ocorre uma forte competição entre mutantes benéficos distintos até que o mutante mais adaptado alcance a fixação com a extinção dos restantes. Este fenômeno é conhecido como Interferência Clonal (Hill and Robertson 1966). A interferência clonal retarda a fixação e conseqüentemente resulta em taxas menores de adaptação em populações assexuadas. O fenômeno da interferência clonal foi extensivamente estudado tanto em uma abordagem teórica (Campos and de Oliveira 2004; de Oliveira and Campos 2004; Orr 2000; Gerrish and Lenski 1998) quanto diversos resultados experimentais foram obtidos (Gerrish and Lenski 1998; Cuevas et al. 2002; Miralles et al. 1999). Estes estudos levaram a uma melhor compreensão dos mecanismos evolucionários que afetam a importância da interferência clonal em populações naturais. Apesar dos avanços, todas as investigações foram restritas à hipótese de populações não estruturadas. Uma primeira tentativa de atacar o problema foi a análise da dinâmica de fixação de mutações vantajosas em uma população haplóide assexuada espacialmente estruturada (Gordo and Campos 2006). Neste modelo, os indivíduos foram distribuídos numa rede bidimensional com competição local, isto é, os organismos interagem somente com seus vizinhos mais próximos. Mostrou-se que a taxa de substituição de mutações benéficas em uma população espacialmente estruturada é menor do que a observada em populações não-estruturada. Esse efeito é mais

intenso no regime de grandes taxas de mutações benéficas, exatamente onde a interferência clonal se torna mais relevante. No entanto, o papel da topologia da rede no processo adaptativo de populações subdivididas ainda não foi analisado. Talvez por causa da hipótese de que a estrutura da população não influencia a dinâmica da população. De fato, Maruyama demonstrou que para determinados tipos de estrutura de população, mais especificamente os modelos onde se supõe migração conservativa, a probabilidade de fixação de mutações apresenta os mesmos resultados daquela verificada para uma população não-dividida (Maruyama 1970a). Isto é precisamente o que observamos em nosso modelo na ausência de extinção e de interferência clonal. Conseqüentemente, esta dissertação concentra-se na competição entre diferentes clones para atingir a fixação, e como este processo de competição afeta a dinâmica da população. Desejamos ainda entender em que condições a topologia das redes migratórias afetam esta competição. Aqui, nós consideramos um modelo de população estruturada introduzido recentemente para investigar variação genética neutra no contexto do modelo epidemiológico suscetível-infectado-suscetível (SIS)(Campos and Gordo 2006). Aqui o modelo é ligeiramente modificado permitindo-se que os indivíduos sofram a ação da seleção natural.

Nos anos recentes uma série de investigações têm sido desenvolvidas para melhor compreender como se influenciam mutuamente as topologias dos sistemas, em nosso caso populações, e a dinâmica desses sistemas. Diversos estudos foram realizados no contexto epidemiológico (Pastor-Satorras and Vespignani 2001; Barthélemy et al. 2004). Verificou-se por exemplo que as doenças se espalham muito mais rapidamente em redes cuja topologia são conhecidas como redes livre de escala (Albert et al. 1999) do que para outros tipo de redes. Essas redes livre de escala foram por exemplo observadas em redes de contato sexuais (Pastor-Satorras and Vespignani 2001). Recentemente, Combadão e colaboradores verificaram que um outro ingrediente deve ser levado em consideração em estudos desse tipo: a dinâmica dos patógenos (Patógeno é um organismo capaz de produzir doença ou dano em um hospedeiro). Verificou-se que em redes livre de escala o acúmulo de mutações deletérias em populações de patógenos é maior, do que por exemplo em grafos aleatórios. Assim, além da própria dinâmica da doença a dinâmica dos patógenos é fundamental para se chegar a uma análise correta do problema.

Nesta dissertação estamos interessados em estudar a dinâmica adaptativa da população e

como a topologia de rede de interação entre as subpopulações afeta o processo evolutivo. Desejamos estudar em que condições aumentamos ou reduzimos a velocidade de adaptação das populações. Aqui, medimos a velocidade de adaptação através da taxa de fixação de mutações benéficas. Damos ênfase à investigação sobre como a migração e o padrão da rede de migrações altera o processo evolutivo, e em que condições a conjectura de Maruyama é válida.

Abaixo realizamos uma breve descrição da organização desta dissertação que está subdividida em 6 capítulos.

No Capítulo 2, apresentamos uma breve discussão sobre a definição de redes e seus padrões de conectividade. Em particular discutiremos as topologias consideradas nesta dissertação, os grafos aleatórios, que foram introduzidos por Erdos e Rényi há aproximadamente 50 anos (Erdos and Rényi 1959), e também apresentamos um modelo de redes recentemente proposto que é o modelo de redes livre de escala (*scale-free networks*), introduzido por Albert e Barabási (Albert, Jeong, and Barabási 1999; Albert and Barabási 2002).

No Capítulo 3, denominado de Dinâmica Adaptativa, definimos o que são mutações e suas classificações. Posteriormente, discutimos como se dá o processo de fixação de mutações benéficas e em que condições estas são importantes. Apresentamos ainda uma abordagem analítica, apresentada por Haldane (Haldane 1927), em que derivamos a probabilidade de fixação para uma mutação benéfica em um modelo de dois alelos.

No Capítulo 4, definimos o que é um modelo de metapopulações, e posteriormente apresentamos o modelo de população estruturada proposto por Campos e Gordo (Campos and Gordo 2006). Neste mesmo capítulo mostramos a correspondência entre esse modelo e o modelo epidemiológico suscetível-infectado-suscetível (modelo SIS). Por fim, introduzimos modificações no modelo original, onde se assumiu seleção neutra. Incorporamos a ação da seleção natural no nosso modelo, permitindo a ocorrência de mutações deletérias e mutações benéficas.

No Capítulo 5 apresentamos então os resultados de simulação de nosso modelo. Verificamos em que condições a probabilidade de fixação de uma dada mutação benéfica não segue a

conjectura de Maruyama, e realizamos na seqüência um estudo da dinâmica adaptativa propriamente dita. Na última seção do capítulo realizamos um estudo complementar da dinâmica das mutações que não alcançam fixação, pois esta análise nos permite compreender melhor como se dá o processo de difusão e fixação e perda de mutações.

E por fim, no Capítulo 6, apresentamos nossas conclusões.

## Redes Complexas

O estudo de redes complexas tem proporcionado grandes avanços no entendimento das propriedades dos diversos tipos de redes de interações observados em distintos campos da ciência, tais como ciências biológicas e sociais. A quantidade abundante de dados disponíveis sobre esses sistemas tem estimulado um número crescente de contribuições neste tópico. Entre os diversos tipos de redes estudados citamos a rede WWW (World Wide Web) (Albert and Barabási 2002), a rede de interação de proteínas (Albert et al. 1999), a rede de atores (Watts and Strogatz 1998), redes sexuais (Pastor-Satorras and Vespignani 2001; Barthélemy, Barrat, Pastor-Satorras, and Vespignani 2004), entre tantos outros.

Por um longo tempo, o estudo de redes complexas ficou restrito à noção de grafos aleatórios, introduzida por Erdős e Rényi (Erdős and Rényi 1959) quase cinquenta anos atrás. Neste modelo, cada nó da rede é conectado a qualquer outro nó com probabilidade  $p$ , produzindo uma rede estatisticamente homogênea, em que a distribuição de conectividade é dada por uma distribuição de Poisson. Uma das principais características da rede resultante é a existência de uma pequena distância entre pares de nós, que é estimada como o número mínimo de vértices entre esses nós. Esta medida topológica apresenta um crescimento logarítmico com o tamanho do sistema. Este fenômeno ficou posteriormente conhecido como efeito de mundo pequeno, motivado principalmente pelo trabalho do psicólogo social S. Milgran que introduziu o conceito de seis graus de separação (Milgran 1967). A observação de distâncias curtas entre nós em redes é um aspecto bem estabelecido de redes de interações reais. Como exemplos podemos citar os resultados experimentais em redes metabólicas (Albert, Jeong, and Barabási 1999), a World-Wide Web (Albert and Barabási 2002), e também a rede de colaboração entre atores (Watts and Strogatz 1998), entre outros.

Embora grafos aleatórios forneçam a ferramenta de modelagem mais prática para a inves-

tigação de redes complexas aleatórias, estudos recentes de sistemas reais asseguram que este modelo paradigmático não é capaz de capturar todas as propriedades topológicas essenciais observadas na maioria dos sistemas. Por esta razão, assumiremos aqui além dos grafos aleatórios como estruturas de nossas populações, as redes livre de escala que são uma proposta mais recente para topologias de alguns sistemas reais. A utilização desses dois tipos de redes nos permitirá entender se existe alguma divergência nos resultados de dinâmica populacional quando modelos padrões e irrealísticos são assumidos, e aqueles mais próximos da realidade.

O estudo das redes, na forma da Teoria dos Grafos é um dos pilares fundamentais da matemática discreta. Atribui-se a Euler a primeira prova neste campo, elaborada em 1735.

## 2.1 Introdução à Teoria dos Grafos

Uma rede complexa é definida como aquela cuja estrutura não segue um padrão regular. No entanto, há dificuldade em se encontrar na literatura um conceito claro e universalmente aceito de padrão regular aplicável a este contexto. Alguns trabalhos entendem redes complexas como aquelas que modelam grandes sistemas, privilegiando o tamanho. No entanto, algumas abordagens impõem um grau de imprecisão aqui indesejado. Na tentativa de eliminar esta imprecisão, a solução encontrada foi fazer um analogia entre redes complexas e grafos e construir a partir da teoria dos grafos, a discussão de interesse.

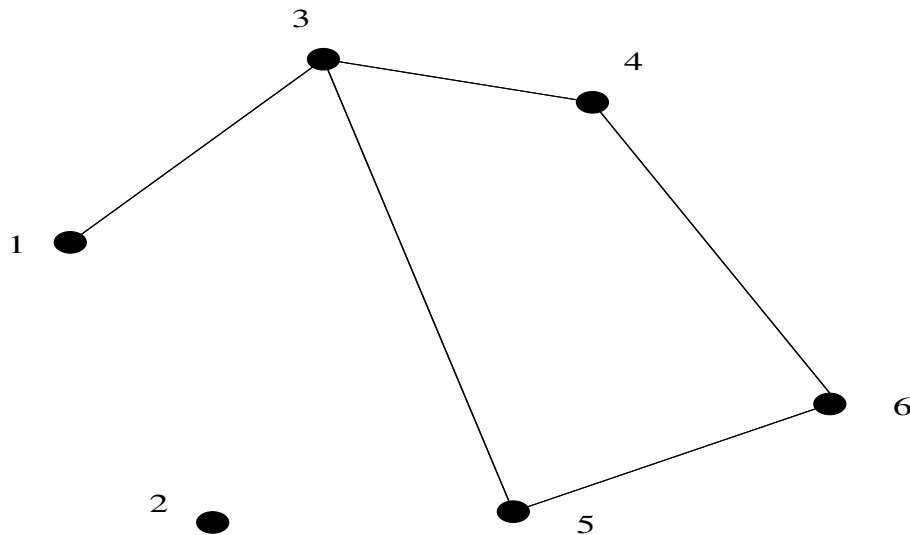
Em termos matemáticos uma rede é representada por um grafo. Um grafo é um par de conjuntos  $G = \{P, E\}$ , onde  $P$  é um conjunto de  $N$  nós ou vértices  $P_1, P_2, \dots, P_N$  e  $E$  é o conjunto de arestas que conectam dois elementos de  $P$ . Os grafos são usualmente representados por conjuntos de pontos, cada um correspondendo a um nó, dois destes nós ligados por uma aresta.

Na Figura 2.1 ilustramos um grafo com  $N = 6$  nós e  $n = 5$  arestas. O conjunto de nós é  $P = \{1, 2, 3, 4, 5, 6\}$  e o conjunto de arestas é  $E = \{(1, 3), (3, 5), (3, 4), (5, 6), (4, 6)\}$ .

As definições abaixo aplicam-se somente a grafos sem laços:(Um laço é a ligação de um nó com ele próprio).

- **ORDEM:** a ordem de grafo  $G$  é dada pelo seu número de nós.





**Figura 2.1** Grafo com  $N = 6$  nós e  $n=5$  arestas. O conjunto de nós é  $P = \{1, 2, 3, 4, 5, 6\}$  e o conjunto de arestas é  $E = \{(1, 3), (3, 5), (3, 4), (5, 6), (4, 6)\}$ .

- **GRAU:** o grau do nó é o número total de ligações (arestas) incidentes em um nó.
- **GRAFO REGULAR:** é aquele em que todo nó tem o mesmo grau. Por exemplo, uma cadeia uni-dimensional ou uma rede quadrada bi-dimensional.
- **GRAFO IRREGULAR:** é aquele que possui nós com diferentes graus.
- **REDES COMPLEXAS:** as redes complexas são definidas como redes cujos nós podem apresentar diferentes graus entre si. As redes regulares são consideradas casos particulares das redes complexas.

## 2.2 Modelo de Erdos e Rényi para grafos aleatórios

No primeiro artigo clássico de Grafos Aleatórios, Erdos e Rényi (Erdos and Rényi 1959) definiram grafos aleatórios com  $N$  nós rotulados conectados por  $n$  arestas que são escolhidas aleatoriamente entre as  $\frac{N(N-1)}{2}$  ligações possíveis entre os pares de nós. No total temos  $\binom{N}{2}$  combinações de grafos com  $N$  nós e  $n$  arestas formando um espaço de probabilidade no qual a formação de cada grafo é equiprovável. Uma definição alternativa de um grafo aleatório é

chamada modelo binomial. Uma outra possibilidade é começarmos com  $N$  nós, e então conectar cada par de nós com probabilidade  $p$ . Por conseguinte, o número total de arestas é uma variável aleatória com valor médio  $E(n) = \frac{pN(N-1)}{2}$ . Se  $G_0$  é um grafo com nós  $P_1, P_2, \dots, P_N$  e  $n$  arestas, a probabilidade de se obter este grafo pelo processo de construção de grafo é  $P(G_0) = p^n (1-p)^{\frac{N(N-1)}{2}-n}$ .

Erdos e Rényi (1959) foram os primeiros a estudar a distribuição do grau máximo em grafo aleatório. A distribuição de grau completa foi obtida por Bollobás (Bollobás 1981). Em um grafo aleatório com probabilidade de conexão  $p$ , o grau  $k$  de um dado nó  $i$  segue uma distribuição binomial com parâmetros  $(N-1)$  e  $p$ ,  $B(N-1, p)$  (B se refere à distribuição binomial):

$$P(k_i = k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}. \quad (2.1)$$

A probabilidade de que  $k$  ligações (arestas) sejam ativadas ao nó  $i$  é justamente  $p^k$ , e a probabilidade da ausência de arestas adicionais é  $(1-p)^{N-1-k}$ . Como existem  $\binom{N-1}{k}$  modos equivalentes de selecionar os pontos terminais para estas  $k$  arestas, este fator deve ser multiplicado para obtermos a probabilidade total.

No limite de grandes redes,  $N \rightarrow \infty$  e assumindo  $Np$  finito, a distribuição acima pode ser aproximada por uma distribuição de Poisson de média  $z = (N-1)p$ , ou seja,

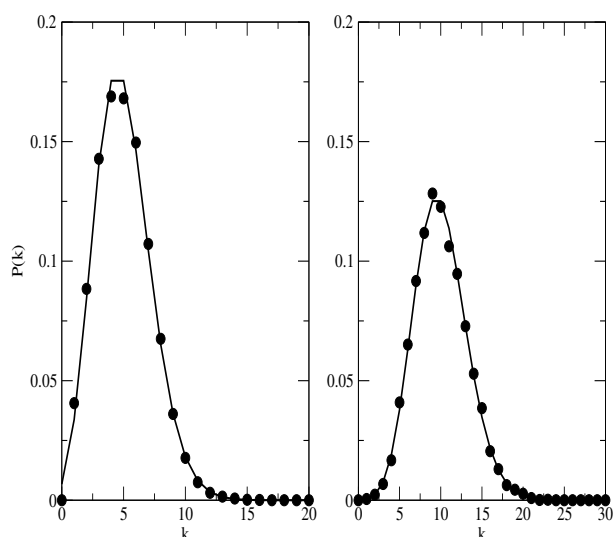
$$P(k_i = k) = \frac{e^{-z} z^k}{k!} \quad (2.2)$$

Neste caso, a variância do grau (conectividade) também é igual a  $z$ .

Na Figura 2.2 exibimos a distribuição de conectividade para uma grafo aleatório com  $N = 10000$  nós e conectividade média  $z = 5$  (Parte (a)) e  $z = 10$  (Parte (b)). Em ambos os casos a distribuição é bem descrita por uma distribuição de Poisson conforme Eq. (2.2). Os pontos são resultados de simulação computacional.

### 2.3 Rede livre de escala

A topologia das redes de diversos sistemas são descritos como redes de topologia complexa. Uma propriedade comum de muitas redes reais é que a conectividade dos vértices segue uma



**Figura 2.2** Distribuição de conectividades de um grafo aleatório com  $N = 10000$  nós e (a) conectividade média  $z = 5$  e (b) conectividade média  $z = 10$ .

distribuição tipo lei de potência, ou seja,  $P(k) \propto k^{-\gamma}$ . Esta característica foi encontrada em diversas redes reais, que variam desde redes de contato sexuais, redes de colaborações científicas, redes de regulação de genes e a rede mundial de computadores (WWW- World Wide Web). A emergência de tal padrão para a distribuição de conectividades é apontada como resultante de dois mecanismos:

- (i) Redes se expandem continuamente pela adição de novos vértices.
- (ii) Vértices novos conectam-se preferencialmente aos vértices que já são bem conectados.

Um modelo baseado nos mecanismos (i) e (ii) reproduz uma distribuição estacionária sem escala definida. Isto indica que o desenvolvimento de redes grandes é governado por uma robusta auto-organização do fenômeno que emerge das particularidades dos sistemas individuais. A inabilidade da ciência contemporânea em descrever sistemas compostos de elementos distintos que interagem entre si, possivelmente a longas distâncias, limita avanços em muitos campos da ciência, variando da biologia molecular até a ciência da computação. A dificuldade para descrever estes sistemas está parcialmente em suas topologias. Muitas destas formam redes bastante complexas cujas arestas representam as interações entre os elementos que compõem o sistema. Por exemplo, a rede de regulação de genes, cujos vértices são proteínas e genes, e

as interações químicas entre eles são representadas por arestas. Em um nível de organização diferente, uma outra rede é formada pelo sistema nervoso, em que os vértices são as células nervosas, conectadas pelos axônios. Redes igualmente complexas aparecem ainda em ciências sociais onde os nós são os indivíduos ou organizações e as arestas são as interações sócio-econômicas entre eles, ou na rede mundial de computadores World Wide Web (WWW), cujos nós são as páginas da WWW (código HTML) conectados por ligações que apontam de uma página para outra. Por causa do grande tamanho e da complexidade das interações, a topologia dessas redes é ainda desconhecida. Tradicionalmente, aplica-se a teoria de grafos aleatórios de Erdos e Rényi (1959) para descrever as redes de topologia complexa. Infelizmente, devido a ausência de dados em redes grandes o suficiente, as previsões desta teoria foram raramente confrontadas com dados obtidos de situações reais. Porém, com aquisição de dados computacionais tais informações sobre as topologias das redes estão cada vez mais disponíveis e assim aumentam as chances de compreensão da dinâmica e estabilidade da topologia de grandes redes.

O primeiro trabalho a demonstrar o surgimento de uma lei de potência para a distribuição de conectividades dos nós de uma rede complexa foi conduzido por Barabási e Albert (Albert, Jeong, and Barabási 1999; Albert and Barabási 2002). Neste trabalho, eles estudaram a topologia das redes WWW, redes de colaboração de atores e também a rede de transmissão de energia elétrica. Todas estas redes foram bem caracterizadas por uma distribuição tipo lei de potência para o grau de conectividade. Este resultado indica que uma grande rede se auto-organiza para um estado livre de escala, um aspecto não previsto por nenhum dos modelos anteriores de redes aleatórias. Os modelos de redes aleatórias assumem uma probabilidade uniforme para que dois nós sejam aleatoriamente conectados. Isto está em desacordo com o observado nas redes reais nas quais certos nós recebem muito mais ligações do que outros. Por exemplo, um novo ator será lançado provavelmente em um papel apoiando atores mais estabelecidos e bem conhecidos. Conseqüentemente a probabilidade de que um novo ator seja lançado contracenando com um ator conhecido é muito mais alta do que o novo ator ser lançado ao lado de atores menos conhecidos.

O surgimento da lei de potência para a distribuição de conectividades observado no modelo de Barabási e Albert (1999) indica que crescimento e ligações preferenciais desempenham um importante papel no desenvolvimento da rede. De forma a verificar que ambos ingredientes são necessários, Barabási e Albert investigaram duas variantes do modelo. Em um modelo que mantém o caráter crescente da rede, mas a ligação preferencial é eliminada, assumindo assim que o novo vértice é conectado com igual probabilidade a qualquer vértice do sistema, conduz o sistema para uma distribuição do tipo  $P(k) \propto e^{-\beta k}$ , indicando que a ausência de ligações preferenciais elimina o aspecto de escala livre da distribuição. Em um segundo modelo, começando com  $N$  nós sem arestas, e a cada passo de tempo nós são então selecionados e conectados com probabilidade

$$P(k_i) = \frac{k_i}{\sum_{j=1}^{N-1} k_j}, \quad (2.3)$$

verificou-se que embora o modelo exiba inicialmente uma lei de potência,  $P(k)$  não é estacionária, pois  $N$  é constante e o número de arestas aumenta com tempo. Após  $T = N^2$  passos de tempo o sistema alcança um estado em que todos os nós são conectados. As falhas dos dois modelos indicam que ambos os ingredientes, crescimento e ligação preferencial, são necessários para o surgimento da distribuição estacionária da lei de potência observada. Por causa da ligação preferencial, um vértice que adquire mais conexões do que outro aumentará sua conectividade numa maior taxa, assim uma diferença inicial na conectividade entre dois vértices aumentará ainda mais com o crescimento da rede. Em uma aproximação contínua, introduzida por Barabási, Albert e Jeong (1999), calcula-se a dependência com o tempo do grau  $k_i$  de um certo nó  $i$ . Este grau aumentará sempre que um novo nó entrar no sistema e se ligará ao nó  $i$  com probabilidade  $P(k_i)$ . Assumindo que  $k_i$  é uma variável real contínua, sua taxa de variação temporal é proporcional a  $P(k_i)$ . Conseqüentemente,  $k_i$  satisfaz a equação dinâmica

$$\frac{dk_i}{dt} = m \frac{k_i}{\sum_{j=1}^{N-1} k_j}$$

Onde no denominador temos a somatório das conectividades de todos os nós do sistema, exceto o nó recém chegado, que é dado por  $\sum_{j=1}^{N-1} k_j = 2mt - m$ . No limite para grandes valores

de  $t$  e substituindo na equação (2.4) obtém-se a equação

$$\frac{dk_i}{dt} = \frac{k_i}{2t}. \quad (2.4)$$

A solução desta equação, com a condição inicial  $k_i(t_i) = m$  é  $k_i(t) = m\sqrt{\frac{t}{t_i}}$ , onde  $t_i$  é o instante em que o nó  $i$  é adicionado ao sistema. Esta é uma propriedade de escalamento dependente do tempo que poderia ser diretamente testada a medida que dados da conectividade da rede tornam-se disponíveis. Assim, antigos nós ( $t_i$  pequeno) aumentam suas conectividades às custas dos nós mais novos ( $t_i$  grande), e nós que já são fortemente conectados, se tornam ainda mais conectados, causando o fenômeno *rich-get-richer* (a água corre sempre para o mar, conforme ditado popular) que pode ser facilmente detectado numa rede real. Esta propriedade pode ser usada para calcular o expoente  $\gamma$  da distribuição de conectividades analiticamente. A probabilidade que um nó  $i$  tenha uma conectividade menor do que  $k$ ,  $P[k_i(t) < k]$ , pode ser escrita como  $P(t_i > \frac{m^2 t}{k^2})$ . Supondo que adicionamos os nós ao sistema em intervalos de tempo iguais, Os valores de  $t_i$  segue uma distribuição uniforme com função de densidade  $f(t_i) = \frac{1}{m_0 + t}$ . Assim podemos escrever:

$$P(t_i > \frac{m^2 t}{k^2}) = 1 - P(t_i \leq \frac{m^2 t}{k^2}) \quad (2.5)$$

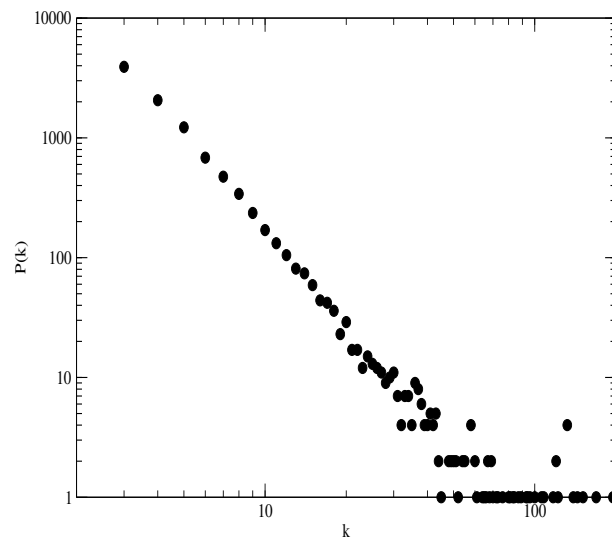
$$= 1 - \frac{m^2 t}{k^2(t + m_0)} \quad (2.6)$$

A densidade de probabilidade  $P(k)$  pode ser obtida por  $P(k) = \frac{\partial P[k_i(t) \leq k]}{\partial k}$  que para períodos longos conduz a solução estacionária.

$$P(k) = \frac{2m^2}{k^3} \quad (2.7)$$

O expoente  $\gamma = 3$  é independente de  $m$ .

A Figura 2.3 apresenta um exemplo da distribuição de conectividades  $P(k)$  para uma rede livre de escala, com  $N = 1000$  nós. A figura está numa escala log-log, e verificamos que a distribuição se aproxima de uma reta, o que comprova que a distribuição é uma lei de potência. A rede foi gerada utilizando-se o algoritmo padrão de Albert-Barabási (Albert, Jeong, and Barabási 1999), que assume crescimento contínuo e ligação preferencial.



**Figura 2.3** Distribuição de conectividades para uma rede livre de com  $N = 10000$  nós. Rede gerada de acordo com o algoritmo de Albert-Barabási. O expoente obtido da distribuição é  $\gamma = 2.7$ .

Antes de iniciarmos nossas aplicações do estudo de redes complexas, vamos realizar no próximo capítulo uma descrição das forças e mecanismos evolucionários que determinam a evolução das populações.





## Dinâmica Adaptativa

Os organismos vivos estão em contínua busca por seu aprimoramento genético que tem por único objetivo permitir uma maior adaptação do ser vivo ao meio em que este reside. O termo adaptação para organismos complexos como seres humanos ou mesmo animais é um tanto difícil de quantificar, mas para microorganismos como bactérias ou vírus a adaptação está estreitamente relacionada às suas taxas de reprodução. Assim, dizemos que uma bactéria que se replica muito rapidamente está muito adaptada ao meio e, portanto, possui um alto valor seletivo (*fitness*). O processo evolutivo se dá pela ação conjunta de duas forças antagônicas: seleção natural e mecanismos de geração de diversidade genética. Entre os mecanismos de geração de diversidade genética destacam-se as mutações. Enquanto a seleção natural tende a reduzir a diversidade genética, com a sobrevivência do mais apto, as mutações geram diversidade. A evolução se dá quando existe um equilíbrio entre essas duas forças antagônicas, já que sem diversidade não há seleção natural.

Em populações finitas, um mecanismo evolucionário se torna muito atuante: esse mecanismo é denominado de deriva genética. A deriva genética também atua no sentido de reduzir a diversidade genética nas populações reais. A deriva genética é um mecanismo que, atuando em consonância com a seleção natural, modifica as características das espécies ao longo do tempo. É um processo estocástico, atuante sobre as populações, modificando as frequências dos alelos e a predominância de certas características na população. É mais freqüente ocorrer em populações com tamanhos reduzidos. A deriva genética e a seleção natural raramente ocorrem independentemente; estes dois fenômenos estão sempre a atuar numa mesma população. No entanto, o grau em que cada alelo é afetado por estes dois fenômenos pode variar em função das circunstâncias. Numa população com tamanho elevado, a deriva ocorre muito lentamente, e a seleção atuante sobre um alelo pode ocorrer de uma maneira relativamente rápida, aumentando ou diminuindo a sua frequência (dependendo da viabilidade do alelo). Numa população

com tamanho reduzido, o efeito da deriva genética predomina. Neste caso o efeito da seleção natural é menos visível pois o efeito da deriva muitas vezes se sobrepõe. A deriva genética pode ter efeitos dramáticos na evolução das populações. Quando ocorre o efeito gargalo (redução significativa no tamanho da população de uma espécie por morte ou por interrupção no processo reprodutivo), a deriva genética pode resultar em rápidas e drásticas alterações nas frequências alélicas. Este fato ocorre de maneira independente da seleção natural. Nestes casos, muitas adaptações benéficas poderão ser eliminadas da população. Pensa-se que o efeito gargalo ocorreu diversas vezes na história da evolução humana. Assim sendo, adiante faremos uma discussão mais detalhada sobre as mutações e suas classificações.

### 3.1 Mutações

Mutação é uma modificação casual ou induzida na informação genética. A mutação só é passada para descendentes de organismos complexos se ocorrer em células germinativas. Para que haja mutação é necessário primeiro que ocorra um dano na seqüência de nucleotídeos do DNA. As células possuem um arsenal de mecanismos de reparação do DNA encarregados de anular o possível dano. Ocasionalmente pode ocorrer uma falha nesses mecanismos (ou o dano é simplesmente irreparável), e as células replicam-se nestas condições. Ainda, as células replicadas com danos no DNA raramente persistem. Apenas uma pequena proporção de células sobrevive carregando os danos genéticos da célula-mãe, passando a apresentar estas novas características. Nestas circunstâncias podemos dizer que ocorreu uma mutação.

As mutações podem ocorrer devido à pequena probabilidade de erro espontâneo no momento da duplicação do DNA, durante a mitose ou meiose. Elas podem ser provocadas por agentes mutagênicos de origem eletromagnética, química ou biológicas; ou ainda podem ser induzidas em laboratório com o uso intencional destes mesmos agentes sobre organismos vivos. As bactérias são ótimos modelos para estudar mecanismos genéticos, por que além de se reproduzirem rapidamente, têm uma cópia de cada gene. Portanto qualquer mutação conduz diretamente a uma alteração fenotípica.

As mutações atuam de forma crucial na evolução das espécies. As alterações morfológicas, nas quais a teoria da Seleção Natural se baseia, se devem a mutações que promovem o surgimento de novas características em determinada população, que por um motivo ou outro faz com que seus portadores sejam mais bem sucedidos que seus concorrentes e predecessores. Da mesma forma, mutações que produzem indivíduos menos adaptáveis ao seu meio tendem a ser rapidamente eliminados por seus concorrentes, já que a probabilidade de um indivíduo menos adaptado reproduzir é menor.

As mutações podem ser classificadas como mutações benéficas, deletérias e neutras.

(i) Mutações benéficas:

Uma percentagem muito pequena de mutações têm um efeito positivo. Estas mutações conduzem a novas versões das proteínas que ajudam um dado organismo, e suas gerações futuras, melhor se adaptar as mudanças em seu ambiente. Por exemplo, a deleção específica do par base 32 em CCR5 humano (CCR5-32) tornam os homozigotos resistentes ao HIV enquanto que retarda o início da AIDS nos heterozigotos. A mutação CCR5 é mais comum nos descendentes europeus. Uma teoria para etiologia da frequência relativamente alta de CCR5-32 na população européia é atribuída à resistência a peste bubônica. Além do fato de que apenas uma fração diminuta das mutações têm efeito positivo, a grande maioria delas conduzem a pequenas variações de vantagem seletiva. Na verdade, as vantagens seletivas das mutações benéficas são distribuída exponencialmente. Além disto, a geração de mutações benéficas não garante sua propagação na população. A grande maioria das mutações benéficas é eliminada por deriva genética, que são efeitos estocásticos decorrentes da finitude das populações. Alguns experimentos demonstram que mudanças no meio no qual uma população está inserida pode levar à uma produção contínua de mutações benéficas. Entretanto, mutações benéficas são também importantes em meios estáveis. Se mutações deletérias de pouco efeito ocorrem regularmente, então é esperado que algumas destas se fixem na população. Este processo levará a um permanente declínio na adaptação dos indivíduos, a menos que mutações reversas ou mutações compensatórias também surjam e se fixem regularmente. Neste contexto, mutações reversas e mutações compensatórias também são consideradas benéficas. Parece

ser bastante provável que alguns desses processos realmente ocorram em populações reais. Pois, existe evidência substancial de que as mutações deletérias de pequeno efeito são bastante comuns (Ohta 1992). Além do mais, as mutações benéficas desempenham papel fundamental no aprimoramento genético de espécies e organismos em geral em meios estáveis.

(ii) Mutações deletérias:

As mudanças no DNA causadas pelas mutações podem causar erros na sequência de proteínas, criando proteínas parcial ou totalmente não-funcionais. Para funcionar corretamente, cada célula depende de milhares de proteínas para funcionar nos lugares certos no tempo certo. Quando uma mutação altera uma proteína que desempenha um papel crítico no corpo, pode resultar uma condição médica. Uma circunstância causada por mutações em um ou mais genes é chamada de desordem genética. Entretanto, somente uma percentagem pequena de mutações causa desordem genética; a maioria não tem nenhum impacto na saúde. Por exemplo, algumas mutações alteram a sequência primária do DNA de um gene mas não mudam a função da proteína fabricada pelo gene. Se uma mutação estiver em uma célula germinal pode fazer com que a prole carregue a mutação em todas as suas células. Este é o caso de doenças hereditárias. Por outro lado, uma mutação pode ocorrer numa célula somática de um organismo. Tais mutações estarão em todos os descendentes desta célula, e determinadas mutações podem fazer com que a célula se torne maligna, e desencadear câncer. Frequentemente, as mutações do gene que poderiam causar desordem genética são reparadas pelo sistema de reparo do DNA da célula. Cada célula tem um certo número de possibilidades em que as enzimas reconhecem e reparam erros do DNA. Como o DNA pode ser danificado ou pode sofrer mutações de muitas maneiras, o processo de reparo do DNA é uma forma importante com que um corpo se protege de doenças.

(iii) Mutações neutras:

Uma mutação neutra é aquela que não altera a adaptabilidade do organismo ao meio em que está inserido. Como exemplo, uma mutação neutra é a que ocorre em um códon de

aminoácido que resulta no uso de um diferente, mas quimicamente similar aminoácido. Isto é resultado do fato da existência de degenerescência na codificação de amino ácidos.

### 3.2 Fixação de mutações benéficas

Adaptação de populações acontece com a ocorrência e subsequente fixação de mutações que conferem alguma vantagem seletiva aos indivíduos que as adquirem. Por esta razão, o estudo de processo de fixação de mutações vantajosas tem uma longa história na literatura de genética de populações. As investigações têm focado em investigar as principais forças evolucionárias que afetam o alcance da seleção natural em populações.

Adaptação, em seu cenário mais simples, consiste de raras mutações benéficas que por meio de rápidas varreduras seletivas alcançam fixação na população. O fato de uma mutação benéfica ter sido gerada, não garante sua fixação na população. Neste processo, a deriva genética como também o fluxo de mutações deletérias desempenham papéis importantes.

O primeiro pesquisador a investigar o processo de fixação de um alelo benéfico em uma população foi Haldane (Haldane 1927). Em seu modelo, ele assumiu a existência de dois tipos de indivíduos:  $A_1$  com taxa de replicação igual a 1, e o indivíduo  $A_2$  que possui o alelo benéfico e que possui taxa de replicação  $1 + s_b$ , onde  $s_b$  é denominada de vantagem seletiva do alelo benéfico. Haldane então investigou a probabilidade de fixação do alelo benéfico quando este surge em um único indivíduo na população.

Para estudar o modelo, Haldane utilizou a formulação de processos de ramificação, que fora aplicado ao estudo probabilístico de extinção de nomes de famílias por Galton e Watson.

### 3.3 Processo de Ramificação de Galton-Watson

Seja  $K$  uma variável aleatória discreta, a função geratriz de probabilidade  $G$  de  $k$  é definida por

$$G_n(z) = \sum_{k=0}^{\infty} p_k z^k, \quad |z| \leq 1, k = 0, 1, 2, \dots \quad (3.1)$$

onde  $p_k$  é a probabilidade de termos  $k$  indivíduos em uma determinada geração  $n$ , e  $\{p_k\}$  define a distribuição de probabilidade do número de indivíduos com a mutação benéfica.

No início consideramos que a população composta por  $N$  indivíduos possui um único indivíduo com a mutação benéfica, o que corresponde a condição inicial  $G_0(z) = z$ . Para obtermos a distribuição  $P(K = k) = p_k$  na geração  $n=1$  usaremos a relação  $G_1(z) = G(z)$ . Generalizando para qualquer  $n$ , temos

$$G_{n+1}(z) = G_n[G(z)] \quad n = 1, 2, 3, \dots \quad (3.2)$$

Se desejarmos encontrar a distribuição do número de indivíduos numa dada geração  $n$  podemos usar a seguinte relação de recorrência da função geratriz de probabilidade.

$$G_{n+1} = G[G_n] \quad n = 1, 2, 3, \dots \quad (3.3)$$

Como desejamos obter a distribuição do número de indivíduos que possuem a mutação benéfica no regime estacionário, teremos  $G_{n+1} = G_n$ .

Cada indivíduo se replica com probabilidade proporcional ao seu valor de adaptação. Sabemos que a população é composta inicialmente por  $N$  indivíduos, e apenas um indivíduo possui a mutação benéfica, e que possui valor de adaptação  $1 + s_b$ . Portanto, a probabilidade que um certo indivíduo na geração  $n = 1$  seja também do tipo  $A_2$  pode ser escrita na forma:

$$p = \frac{1 + s_b}{N + s_b} \quad (3.4)$$

Já que  $N + s_b$  representa a adaptação de toda população, a probabilidade de que o indivíduo mais adaptado seja selecionado  $k$  vezes é dado pela distribuição binomial

$$p_k = \binom{N}{k} p^k (1 - p)^{N-k}, \quad k = 0, 2, 3, \dots, N \quad (3.5)$$

que para valores grandes de  $N$ , o que implica pequenos valores de  $p$ , pode-se aproximar a distribuição binomial por uma distribuição de Poisson com média  $Np \simeq 1 + s_b$ , e logo podemos escrever

$$p_k = \frac{e^{-(1+s_b)} (1 + s_b)^k}{k!} \quad (3.6)$$

Substituindo a Eq. (3.6) na Eq. (3.1) temos

$$G(z) = \sum_{k=0}^{\infty} \frac{e^{-(1+s_b)} (1+s_b)^k z^k}{k!} \quad (3.7)$$

$$= e^{-(1+s_b)} \sum_{k=0}^{\infty} \frac{((1+s_b)z)^k}{k!} \quad (3.8)$$

$$= e^{-(1+s_b)} e^{(1+s_b)z} \quad (3.9)$$

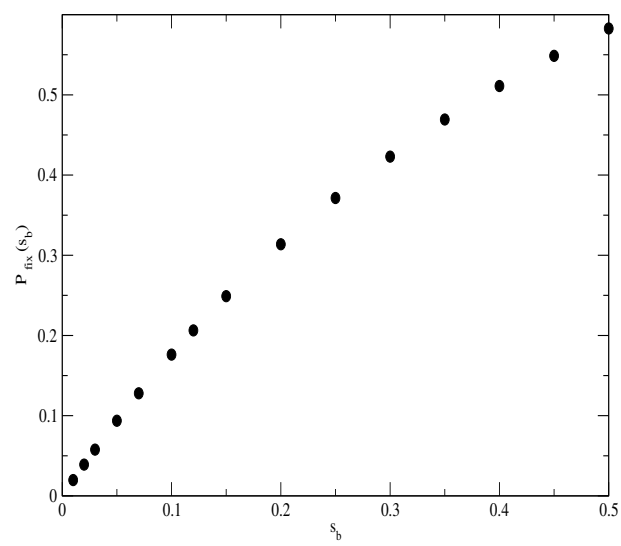
$$= e^{-(1+s_b)(1-z)} \quad (3.10)$$

Como desejamos a solução de regime estacionário, teremos  $G(z) = z$ . Portanto  $z = e^{-(1+s_b)(1-z)}$  e a probabilidade de fixação  $P_{fix}$  que é  $P_{fix} = 1 - z$  pode ser obtida da relação

$$P_{fix} = 1 - e^{-(1+s_b)P_{fix}} \quad (3.11)$$

Mediante uma expansão em série de Taylor para pequenos valores de  $s_b$  até segunda ordem, temos que  $\pi \simeq 2s_b$ . Daí observamos que a probabilidade de fixação de uma mutação benéfica é aproximadamente o dobro do valor da vantagem seletiva que esta confere. Verifica-se experimentalmente que a grande maioria das mutações benéficas conferem baixas vantagens seletivas, e com isso elas raramente atingem a fixação porque são eliminadas devido à ação da deriva genética.

Na Figura 3.1 mostramos a solução da Eq. (3.11) para vários valores de  $s_b$ . Como esperado a probabilidade de fixação  $P_{fix}$  é uma função crescente de  $s_b$ . Podemos observar que, por exemplo, mesmo para um valor alto de vantagem seletiva  $s_b = 0.5$  a probabilidade de fixação é de aproximadamente 0.6.



**Figura 3.1** Probabilidade de fixação  $P_{fix}$  como função da vantagem seletiva  $s_b$ .



## Modelo de Metapopulações

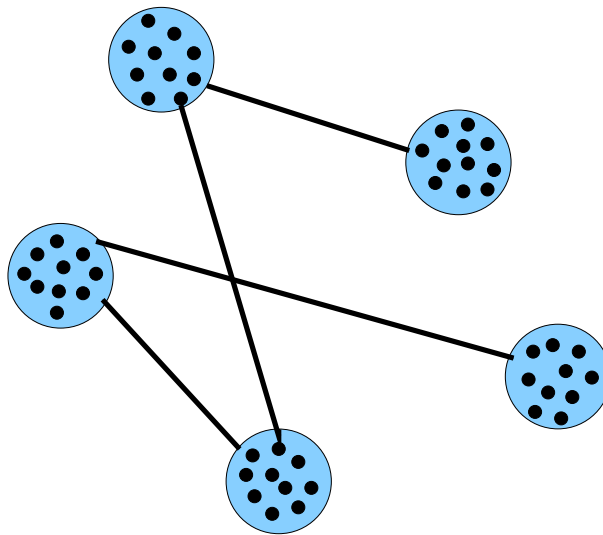
Uma metapopulação consiste de um grupo de subpopulações de mesma espécie separadas espacialmente e que possuem um certo grau de interação. O termo metapopulação foi criado por Richard Levins em 1969 (Levins 1969) para descrever um modelo de dinâmica de população da peste de insetos em campos de agricultura. A idéia foi mais amplamente aplicada dentro de habitats naturais ou artificialmente fragmentados. Geralmente, considera-se uma metapopulação constituída de várias subpopulações distintas juntas com áreas de habitat adequadas que são normalmente desocupadas. Cada subpopulação evolui com relativa independência das outras subpopulações e eventualmente vai à extinção como consequência da estocasticidade demográfica (flutuações no tamanho da subpopulação devido à eventos demográficos aleatórios). Uma pequena população é mais propensa à extinção. Embora as subpopulações tenham tempos de vida finitos, a metapopulação é freqüentemente estável devido à migração de indivíduos entre as subpopulações. Onde os imigrantes serão os prováveis recolonizadores do habitat que foi deixado pela extinção de uma outra subpopulação. O desenvolvimento da teoria de metapopulação, conjuntamente com o desenvolvimento da dinâmica nascimento-extinção, enfatizou a importância da conectividade entre as subpopulações aparentemente isoladas. Embora nenhuma subpopulação possa garantir sua sobrevivência de uma dada espécie a longo prazo, o efeito combinado de muitas subpopulações podem fazer isto.

O modelo de metapopulações atraiu pesquisadores de diferentes campos, tais como ecologia, epidemiologia, e também da área de genética de populações, entre tantos outros. Uma recente comparação entre modelos de metapopulações e modelos epidemiológicos revelou enormes semelhanças. Uma conclusão básica que pode ser deduzida desses estudos consiste em que as duas classes de modelos diferem essencialmente na quantidade de detalhes que estão incluídos (Dobson 2003). Do lado epidemiológico, o foco é modelar populações de micróbios que podem causar doenças humanas importantes tais como a malária e a gripe. No contexto da ge-

nética de populações, uma questão crítica é compreender as forças evolucionárias responsáveis pelos padrões de variação genética em populações naturais. Um interesse crescente da Física neste campo foi observado (Sella and Hirsh 2005; Simon and Derrida 2006). A importância em considerar tanto a genética quanto a epidemiologia para modelar doenças infecciosas é cada vez mais reconhecida (Grenfell et al. 2004). A compreensão da variabilidade genética nos patógenos é da maior importância já que pode ajudar a localizar genes sob pressão da seleção criada pelo sistema imunológico (Connway and et al. 2000) e, sob determinadas circunstâncias, pode ser usada para inferir a história evolutiva da população ancestral (Falush and et al. 2003). Além disso, pode refletir o potencial evolucionário da população (McDonald and Linde 2003). Os modelos padrões de genética de população estruturada fazem suposições topológicas muito simples. Os modelos mais estudados são o modelo de ilhas de Wright (Wright 1931a) e o modelo *stepping-stone* de Kimura (Kimura 1953). Nestes modelos, uma metapopulação (população grande) é composta de subpopulações pequenas denominadas demes. Ao aplicar estes modelos à variação do patógeno em estudo, cada deme corresponde a um hospedeiro (indivíduo). No modelo de ilhas, supõe-se que cada deme interaja com todas as outras demes restantes da população. Por outro lado, no modelo *stepping-stone* as interações entre as demes são locais, ou seja um indivíduo interage apenas com seus vizinhos mais próximos.

Nenhum dos dois modelos, nem o de ilhas nem o modelo *stepping-stone*, que pressupõem redes regulares, são descrições exatas da interação em redes de populações reais. Particularmente, em um contexto epidemiológico, esperamos que as interações se assemelhem às das redes sociais (Keeling 2005). Com o grande desenvolvimento na potencialidade de processamento de dados, várias investigações recentes examinaram as topologias de muitos sistemas reais (Newman 2003). De fato, as descobertas destes estudos mostraram que as topologias reais estão longe de serem regulares, tais como o modelo *stepping-stone* ou o modelo de ilhas. Alguns modelos de rede recentemente propostos capturam características essenciais das redes reais, tais como as redes de pequeno-mundo (Watts and Strogatz 1998) e as redes livre de escala (Albert et al. 1999), que são normalmente consideradas como redes complexas.

A formulação de novos modelos de população estruturada provocou um crescente interesse da comunidade científica em compreender a influência mútua entre a topologia das populações



**Figura 4.1** Ilustração de uma população estruturada. Na figura temos uma população com  $D = 5$  demes, e cada deme possui  $N_d = 10$  indivíduos.

e a dinâmica intrínseca dos sistemas. Neste contexto, o estudo do espalhamento de doenças recebeu atenção especial. Por exemplo, Pastor-Satorras e outros (Pastor-Satorras and Vespignani 2001; Barthélemy, Barrat, Pastor-Satorras, and Vespignani 2004) mostraram em um estudo do clássico modelo epidemiológico suscetível-infectado-suscetível (modelo SIS), que as redes livre de escala são as mais indicadas para representar o espalhamento de doenças do que grafos aleatórios e redes regulares. Entretanto, nenhum desses estudos levou em consideração, em sua formulação, o papel da evolução do micróbio no caso de expansão epidêmica, e a interação dessa dinâmica evolutiva com as topologias das redes. Recentemente, Campos e colaboradores (Campos et al. 2006) demonstraram que este é um elemento a ser considerado pois a topologia pode afetar extremamente a taxa de fixação de mutações deletérias e conseqüentemente o potencial evolucionário das populações (Campos et al. 2006).

Recentemente, Campos e Gordo (2006) propuseram modelos de genética de população estruturada, que incorporam explicitamente parâmetros epidemiológicos, a fim de estudar a variabilidade genética do patógeno sob a dinâmica do modelo SIS. O SIS é um dos modelos clássicos mais simples na epidemiologia, e usado geralmente no estudo das doenças sexualmente transmissíveis (Lloyd and R.M. 2001). Basicamente, o modelo de Campos e Gordo (2006) visou investigar, à luz do modelo SIS, os níveis e os testes padrões da variação da seqüência

do patógeno. É este modelo que será utilizado em nosso estudo. Entretanto, o modelo será modificado de forma a incorporar a ocorrência de mutações benéficas.

### 4.1 O modelo Original de Campos e Gordo

Nesta seção descreveremos o modelo que aqui estudaremos. Nós consideramos uma metapopulação de indivíduos (demes). Cada deme pode carregar no máximo  $N_d$  patógenos, onde cada patógeno é representado por um sequência de nucleotídeos infinita  $S = (s_1, s_2, \dots, s_\infty)$ , onde o nucleotídeo  $s_\alpha$  pode assumir dois valores distintos  $s_\alpha = 0$  (estado original) ou  $s_\alpha = 1$  (uma mutação ocorreu). Cada nó da rede corresponde a uma deme, e o número total de demes é  $D$ . Então o número máximo de patógenos na metapopulação é  $N_t = DN_d$ . Na Figura 4.1 ilustramos um exemplo de uma metapopulação com  $D = 5$  demes e  $N_d = 10$  indivíduos dentro de cada deme. Em cada deme pode ocorrer extinção dos patógenos que carrega com probabilidade  $e$  em cada geração. Quando isto ocorre a deme torna-se vazia. No entanto, estamos supondo também a migração, logo as demes vazias podem ser recolonizadas por migração. Esta última etapa faz nosso modelo diferente dos modelos anteriores de metapopulação onde a migração e a recolonização são processos distintos (Pannell and Charlesworth 2000). Em cada geração o número  $n_e$  de emigrantes de uma certa deme  $j$  (que não está vazia) segue uma distribuição de Poisson com média  $N_d m k_j$ , isto é,

$$P(n_e) = \frac{e^{-N_d m k_j} (N_d m k_j)^{n_e}}{n_e!} \quad (4.1)$$

onde  $m$  é taxa de migração e  $k_j$  é a conectividade da deme  $j$ . Os  $n_e$  patógenos são escolhidos aleatoriamente da deme  $j$  e são então distribuídos de forma equiprovável entre as  $k_j$  demes receptoras. Esta suposição pode servir como uma aproximação ao que de fato pode ocorrer na propagação da malária, onde o patógeno é transmitido a um dado indivíduo através da picada do mosquito transmissor. Após a transmissão, a população de patógenos dentro da deme cresce e pode evoluir dentro da deme. Cada patógeno adquire novas mutações de acordo com uma distribuição de Poisson de média  $U$ . Como no modelo original assume-se seleção neutra. Isto é, todos os patógenos contribuem de forma equiprovável para a geração seguinte. Após cada

ciclo de seleção, cada deme não-vazia tem exatamente  $N_d$  patógenos.

## 4.2 Correspondência entre o modelo de metapopulações e o modelo SIS

Vamos agora examinar a correspondência entre o modelo de metapopulação acima e o modelo SIS. No modelo Suscetível-Infetado-Suscetível, os indivíduos podem estar em dois estados: suscetível ou infectado. Um indivíduo suscetível pode tornar-se infectado quando em contato com indivíduos infectados. A transmissão ocorre a uma taxa  $\beta$ . Por outro lado, os indivíduos infectados retornam à classe suscetível a uma taxa  $\alpha$ . No limite determinístico, a evolução do sistema é descrita pelo sistema de equações diferenciais:

$$\frac{dS}{dT} = -\beta SI + \alpha I \quad (4.2)$$

$$\frac{dI}{dT} = \beta SI - \alpha I \quad (4.3)$$

onde  $S$  é a fração de indivíduos suscetíveis e  $I$  corresponde à fração de indivíduos infectados. Todavia, medindo o tempo em unidades de duração da infecção  $t = \alpha T$  e sabendo que  $S + I = 1$ , a evolução da população infectada com o tempo é obtida através da equação

$$\frac{dI}{dt} = R_0 I(1 - I) - I \quad (4.4)$$

onde  $R_0 = \frac{\beta}{\alpha}$  é o parâmetro relevante do modelo e é conhecido como número básico reprodutivo. O parâmetro  $R_0$  de uma infecção representa o número médio de casos secundários que um único indivíduo infectado típico pode infectar em uma população totalmente suscetível. As soluções da equação acima são  $I = 0$  e  $I = 1 - \frac{1}{R_0}$ . A solução  $I = 0$  é estável somente quando  $R_0 < 1$ , enquanto a solução  $I = 1 - \frac{1}{R_0}$  é estável para  $R_0 > 1$ .

Como consideramos em nosso modelo que um indivíduo é representado por uma deme, uma deme vazia corresponde a um indivíduo que está no estado suscetível, enquanto que uma deme cheia corresponde a um indivíduo infectado. Uma deme que está atualmente cheia pode ficar vazia com probabilidade  $e$ . Isto quer dizer que  $e$  corresponde ao parâmetro  $\alpha$  do modelo SIS. Uma deme vazia pode ficar cheia por meio dos migrantes que esta recebe de outras demes.

Assim  $\beta$  é proporcional a  $m$ . Desde que o número de migrantes que uma deme com conectividade  $k_j$  é igual a  $N_d m k_j$ , então  $\beta$  corresponde a  $N_d m K$ , onde  $K$  é a conectividade média da rede.

### 4.3 O modelo modificado com mutações benéficas e deletérias

No modelo original, Gordo e Campos assumiram seleção neutra, ou seja, as mutações não apresentam nenhuma vantagem ou desvantagem seletiva. Aqui nós consideraremos que a seleção natural atua. Assumiremos então a existência tanto de mutações benéficas, que são aquelas que conferem vantagem seletiva aos indivíduos que as adquirem, quanto de mutações deletérias que reduzem a adaptação dos organismos. As mutações deletérias têm efeito constante e cada mutação reduz a adaptação dos organismos por um fator  $(1 - s_d)$ . Além do mais estas mutações ocorrem a uma taxa  $U_d$ , e o número adicional de mutações que uma prole adquire no processo reprodutivo é uma variável aleatória que segue uma distribuição de poisson de modo que

$$P(n) = \frac{e^{-U_d} U_d^n}{n!}. \quad (4.5)$$

Já o efeito seletivo das mutações benéficas não é constante como no caso de mutações deletérias. Cada mutação benéfica aumenta a adaptação dos organismos por um fator  $(1 + s_b)$ , onde  $s_b$  é exponencialmente distribuído conforme

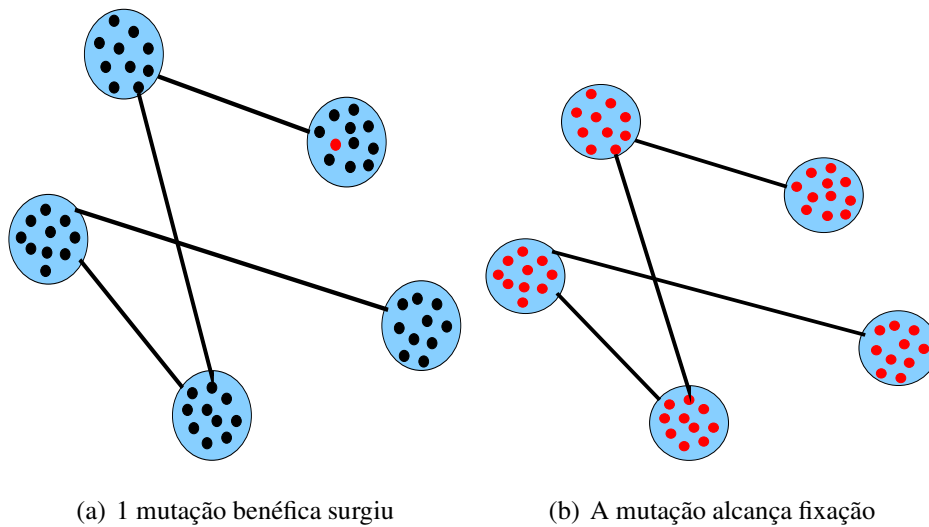
$$P(s_b) = \beta e^{-\beta s_b} \quad (4.6)$$

onde  $1/\beta$  é o valor seletivo médio das mutações benéficas (Gillespie 1991). As mutações benéficas ocorrem a uma taxa constante  $U_b$ , e o número de mutações benéficas que um indivíduo adquire no processo é dado por

$$P(n_b) = \frac{e^{-U_b} U_b^{n_b}}{n_b!}. \quad (4.7)$$

Um indivíduo com  $k_b$  mutações benéficas e  $k_d$  mutações deletérias terá valor de adaptação igual a

$$\omega = \left[ \prod_{i=1}^{k_b} (1 + s_b(i)) \right] (1 - s_d)^{k_d} \quad (4.8)$$



**Figura 4.2** Ilustração do processo de fixação de uma mutação benéfica, aqui representada pela cor vermelha.

A fixação de uma dada mutação ocorre quando todos os indivíduos da população adquirem aquela mutação, ou seja, o indivíduo que primeiro adquiriu aquela mutação se torna ancestral comum da população. Na Figura 4.2 ilustramos o processo de fixação de uma mutação quando esta surge em um indivíduo e se difunde pela população por meio da seleção natural e também por migração.

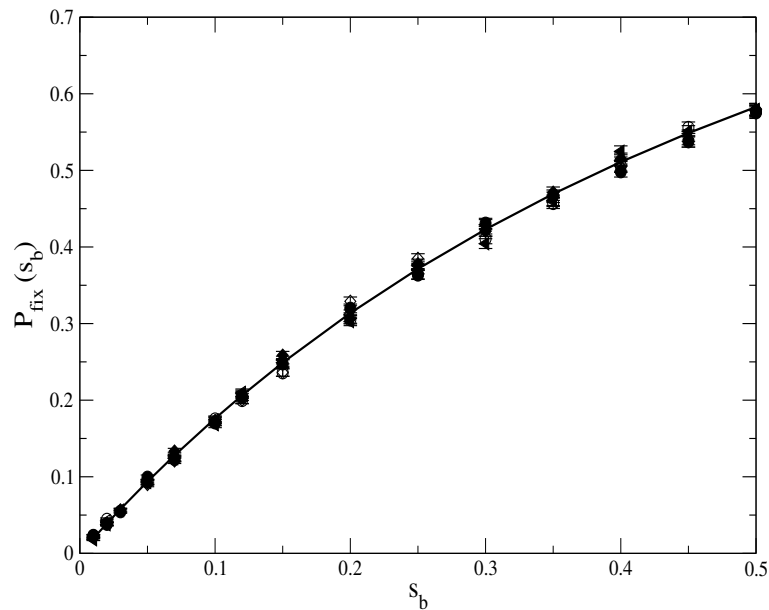




## Resultados e Discussões

Em nosso modelo a migração é dita conservativa, isto é, o número esperado de emigrantes é igual ao número de imigrantes em cada deme. Neste caso específico, Maruyama (1970a) demonstrou que a probabilidade de fixação não depende da estrutura da população, e neste caso a probabilidade de fixação é igual a obtida para uma população homogênea. A fim de testar esta hipótese em nosso modelo, calculamos a probabilidade de fixação das mutações vantajosas em função de seu efeito benéfico  $s_b$ . Iniciamos a nossa população com  $N - 1$  indivíduos com valor de adaptação igual a 1 e um único indivíduo com adaptação  $1 + s_b$  (aqui o valor de  $s_b$  é fixo, em vez de distribuído exponencialmente). Então verificamos se a mutação se fixou na população, ou seja, todos indivíduos adquiriram aquela mutação, ou se a mutação se perdeu. Devemos lembrar o fato de que só existe uma única mutação benéfica e neste estágio não existe nenhuma competição entre mutantes benéficos. Vamos assumir também que a taxa de extinção é igual a zero,  $e = 0$ . Isto corresponde ao conhecido modelo de dois alelos (Haldane 1927). Na Figura 5.1 mostramos a probabilidade de fixação (o número de eventos onde a fixação foi conseguida dividido pelo número de amostras) em função de  $s_b$  para diversas taxas de migração e diferentes topologias (grafo aleatório e Rede livre de escala). Na Figura também comparamos nossos resultados com a predição teórica, dada na Eq. (3.11), para uma população homogênea. O colapso de todas as curvas, demonstra que a suposição de Maruyama (1970a) é realmente válida em nosso modelo.

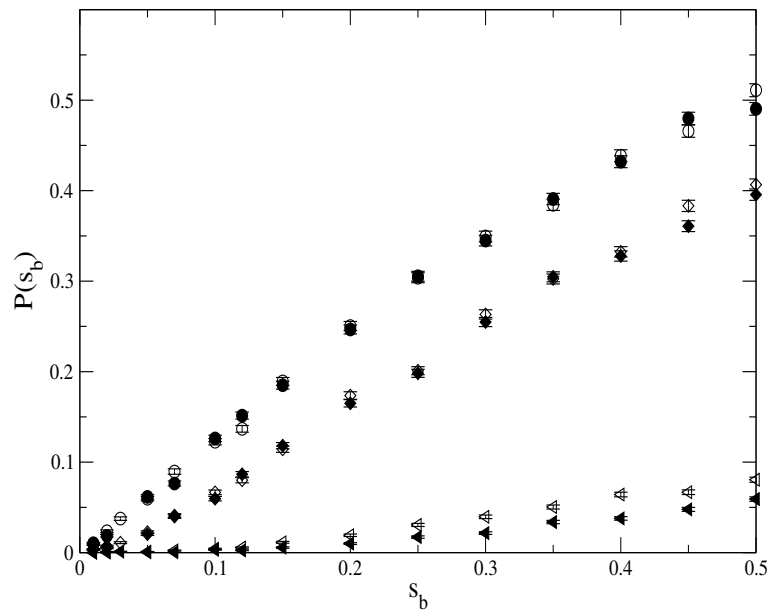
Por outro lado, na Figura 5.2 mostramos a probabilidade de fixação como função de  $s_b$  mas agora considerando-se uma probabilidade de extinção de demes não-nula. Agora vemos claramente a dependência da probabilidade de fixação com a taxa de migração  $m$ . Também observamos uma dependência topológica, especialmente para baixas taxas de migração. Nessas condições, a probabilidade de fixação numa rede livre de escala é ligeiramente maior do que aquela obtida num grafo aleatório. Observamos também que a probabilidade de fixação



**Figura 5.1** Probabilidade de fixação  $P_{fix}$  como função do valor benéfico  $s_b$  da mutação benéfica. Os dados correspondem a resultados para 5000 simulações independentes. Os valores de parâmetro são  $N = 10000$ ,  $D = 200$ ,  $e = 0$ , e  $m = 0.0001$  (triângulos à esquerda),  $m = 0.001$  (diamantes) e  $m = 0.01$  (círculos). Os símbolos cheios correspondem aos resultados para topologias de grafo aleatório, enquanto que os símbolos vazios se referem a topologias do tipo livre de escala. Em ambos os casos a conectividade média é  $z = 6$ . A linha cheia é predição teórica conforme Eq. (3.11). As barras de erro dos valores estimados via simulação são aproximadamente do mesmo tamanho que os símbolos utilizados.

decrece com a diminuição da taxa de migração, ou seja, quanto menor for a taxa de migração  $m$  menor será a probabilidade de fixação de uma dada mutação benéfica. Para taxas de migração muito pequenas, as mutações benéficas têm uma chance maior de eliminação por causa da extinção da deme antes da ocorrência de eventos de migração. Por esta razão a probabilidade de fixação decai com a migração quando a extinção é considerada.

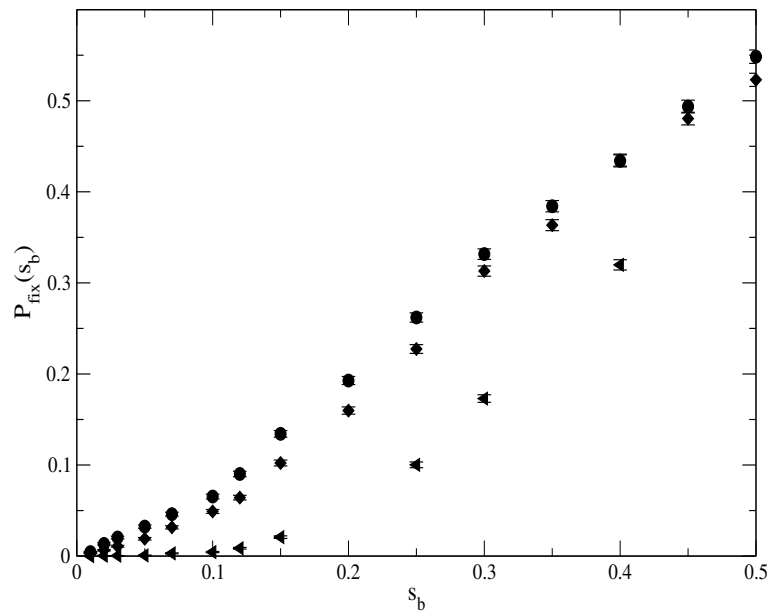
Interessantemente, na Figura 5.3 onde exibimos a probabilidade de fixação como função de  $s_b$  mas agora assumindo a ocorrência de mutações deletérias a uma taxa  $U_d = 0.1$ , vemos que surge uma dependência da probabilidade de fixação com a taxa de migração  $m$ . Este resultado não é previsto por Maruyama. Ressaltando que nas simulações da figura consideramos taxa de extinção nula. A redução maior na probabilidade de fixação é vista para pequenos valores de  $m$ . Como nestes casos, o tempo para espalhamento e difusão de uma dada mutação benéfica



**Figura 5.2** Probabilidade de fixação  $P_{fix}$  como função do valor benéfico  $s_b$  da mutação benéfica. Os parâmetros são os mesmos da Figura 5.1, exceto que  $e = 0.02$ .

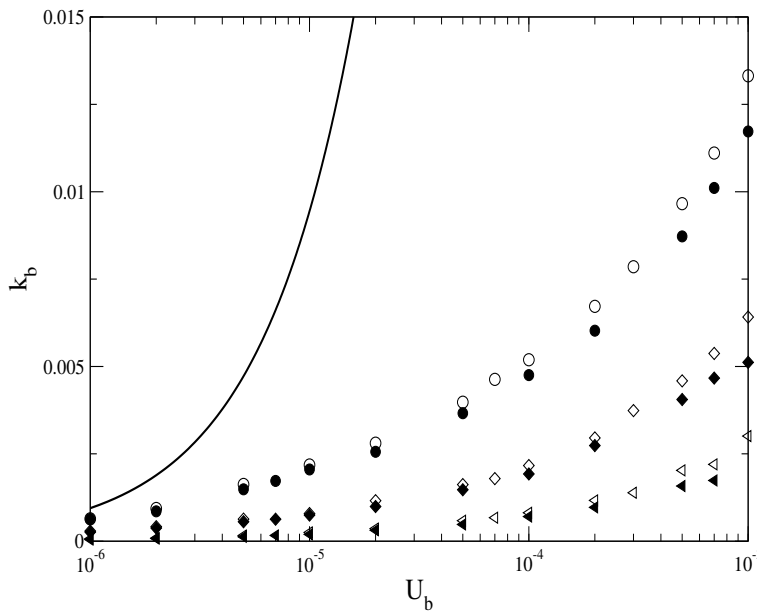
é grande, desde que este tempo é proporcional a  $1/m$ , isto permite a ocorrência de múltiplas mutações deletérias nos indivíduos carregando a mutação benéfica, o que reduz sua adaptação e portanto sua chance de fixação.

Nenhuma das simulações anteriores concebe a competição entre mutações vantajosas. Conseqüentemente nesta análise não é possível apreciar como outros fatores evolucionários, como por exemplo a interferência clonal, podem afetar a dinâmica evolucionária. De agora em diante, suporemos que as mutações benéficas ocorrem a uma taxa constante  $U_b$  e assumiremos  $U_d = 0$ . Iniciamos as populações com todos os indivíduos sem mutações. Deixaremos a população evoluir até a ocorrência de  $N_{ev} = 20$  eventos de fixação e então estimaremos o número de gerações necessárias para a obtenção desses  $N_{ev}$  eventos. O processo é repetido em  $N_{runs}$  independentes a fim de se obter uma estatística razoável para nossos dados de simulação. Verificamos que  $N_{runs} = 50$  fornece uma boa estatística para os dados, ou seja, para este tamanho de amostragem é possível ter uma visão clara do comportamento da taxa de ocorrência de eventos de fixação. Além do mais, uma amostragem maior apresenta um custo computacional extremamente elevado.



**Figura 5.3** Probabilidade de fixação  $P_{fix}$  como função do valor benéfico  $s_b$  da mutação benéfica. Os valores de parâmetro são  $N = 10000$ ,  $D = 200$ ,  $e = 0$ , e  $m = 0.0001$  (triângulos à esquerda),  $m = 0.001$  (diamantes) e  $m = 0.01$  (círculos). Os dados são para topologias do tipo grafo aleatório. A conectividade média é  $z = 6$ , e aqui consideramos uma taxa de mutação deletérias não-nula,  $U_d = 0.1$ .

Na Figura 5.4 exibimos um gráfico da taxa de fixação de mutações benéficas  $k_b$  em função da taxa de mutações benéficas  $U_b$ , para alguns valores de  $m$  e para as duas topologias em questão, os grafos aleatórios e as redes livre de escala. Nesta figura  $e = 0$  e portanto não há extinção das demes. Como esperado, observamos claramente que a taxa de fixação das mutações benéficas cresce com o aumento de  $U_b$ , e que  $k_b$  é fortemente dependente da taxa de migração  $m$ . Uma taxa menor de migração significa efetivamente uma taxa menor de fixação  $k_b$ . Embora a probabilidade de fixação de uma única mutação benéfica seja a mesma para todos os valores da taxa de migração, como mostrado na Figura 5.1, seu tempo para fixação é sensivelmente dependente de  $m$ . Para  $m$  grande a mutação pode rapidamente espalhar-se por toda a rede, enquanto que para  $m$  pequeno o processo de difusão pela população é muito lento. Observamos também que a taxa de crescimento de  $k_b$  diminui enquanto consideramos uma menor taxa de migração. Na Figura 5.4, observa-se uma dependência de  $k_b$  com a topologia da rede, sendo  $k_b$  maior para redes livre de escala. Entretanto, as diferenças entre as duas

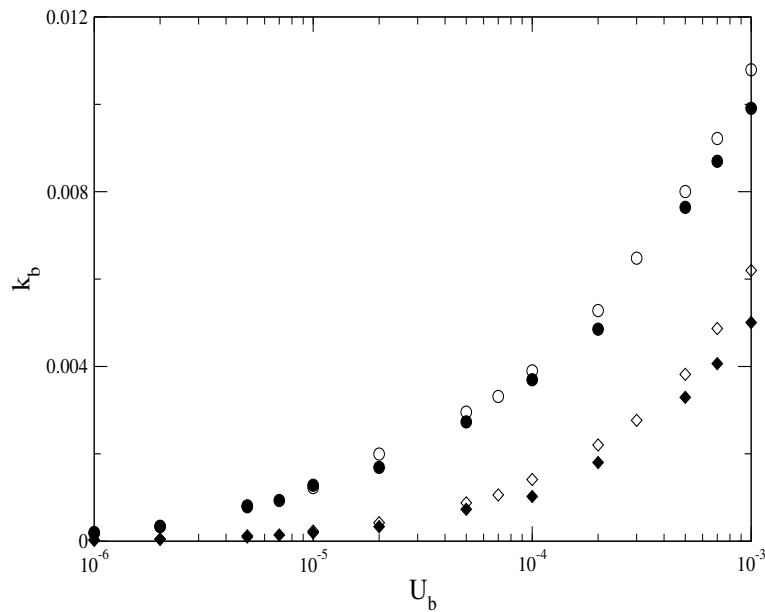


**Figura 5.4** Taxa de fixação de mutações benéficas,  $k_b$ , como função da taxa de mutações benéficas  $U_b$ . Os valores de parâmetro são  $N = 10000$ ,  $D = 200$ ,  $e = 0$ ,  $\beta = 20$ , e  $m = 0.0001$  (triângulos á esquerda),  $m = 0.001$  (diamantes) e  $m = 0.01$  (círculos). Os símbolos cheios correspondem aos resultados para topologias de grafo aleatório, enquanto que os símbolos vazios se referem a topologias do tipo livre de escala. Em ambos os casos a conectividade média é  $z = 6$ . A linha cheia corresponde à taxa esperada para uma população homogênea sem interferência clonal conforme Eq.(5.1).

topologias são destacadas somente para grandes valores de  $U_b$ . Neste regime, um fluxo maior de mutações resulta em uma competição forte entre as diversas mutações, ocorrendo o que chamamos de interferência clonal. Conseqüentemente a fixação de uma dada mutação no meio se deve à eliminação de todos os outros competidores. Junto com os resultados da simulação, mostramos também a taxa de fixação prevista em uma população homogênea infinitamente grande desprezando o efeito da interferência clonal, que é dada por

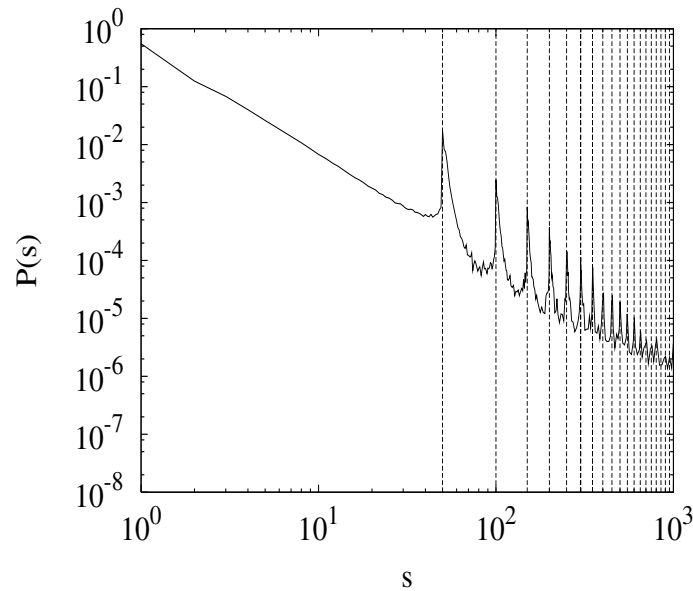
$$K_b = NU_b P_{fix}(\beta), \quad (5.1)$$

onde  $NU_b$  é o número de mutações benéficas por geração e  $P_{fix}(\beta)$  é a probabilidade de fixação de uma dada mutação benéfica cujo efeito seletivo é distribuído exponencialmente (Gillespie 1991). Para  $U_b$  pequeno, onde a interferência clonal não é tão significativa, a taxa de fixação para populações homogêneas não é muito diferente de nossos resultados obtidos da simula-



**Figura 5.5** Taxa de fixação de mutações benéficas,  $k_b$ , como função da taxa de mutações benéficas  $U_b$ . Os parâmetros são os mesmos da Figura 5.4, exceto que  $e = 0.02$ .

ção para populações estruturadas. Todavia, com o crescimento de  $U_b$  a competição entre as mutações benéficas torna-se comum, e a Equação 5.1 passa a superestimar os resultados da simulação. Na Figura 5.5 mostramos a taxa de fixação  $k_b$  como função da taxa de mutação  $U_b$  levando em conta uma taxa de extinção não nula,  $e = 0.02$ . Neste caso, para este valor específico da taxa de extinção, para cada  $n = 50$  gerações, em média, cada deme atravessa um processo de extinção, ou seja a deme torna-se vazia. Após um processo de extinção a deme pode ser recolonizada por migrantes de demes vizinhas. Na Figura, observamos o mesmo cenário qualitativo registrado na figura anterior. Como no caso anterior (sem extinção), a diferença entre as taxas de migração para grafos aleatórios e redes livre de escala é claramente observada no regime com interferência clonal. Em todos os casos, quando a extinção é incluída, a taxa de fixação é reduzida.

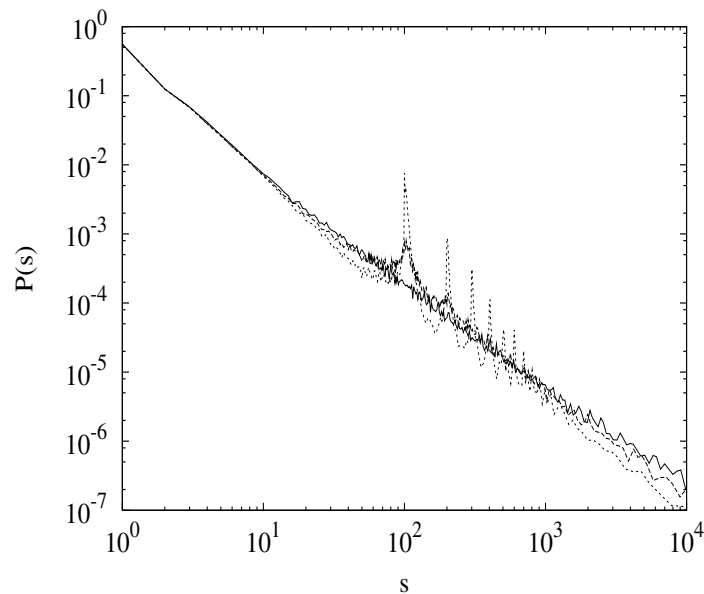


**Figura 5.6** Distribuição de probabilidades de tamanhos de população máximos alcançado pelas mutações que foram posteriormente eliminadas. Os parâmetros são  $N = 10000$ ,  $D = 200$ ,  $z = 4$ ,  $m = 0.0001$  e  $e = 0$ . As linhas verticais indicam múltiplos inteiros do tamanho da deme. A topologia da rede é do tipo grafo aleatório.

## 5.1 Dinâmica das mutações que não alcançam fixação

Recentemente, Rosas e colaboradores (Rosas et al. 2005) investigaram a dinâmica da eliminação de mutações favoráveis numa população assexuada. Eles focaram seu estudo na análise da distribuição de probabilidade  $P(s)$  das mutações atingirem um tamanho máximo de população  $s$  e posteriormente serem extintas devido à ação da deriva genética ou da interferência clonal. Curiosamente, eles mostraram o aparecimento de dois regimes de escala distintos na análise de  $P(s)$ : o primeiro para pequenos tamanhos  $s$ , onde a deriva genética é a principal força evolucionária e a interferência clonal tem um papel sutil, e um segundo regime para  $s$  grande onde a interferência clonal prevalece sobre a deriva genética. Esta análise é particularmente elucidativa por que permite a determinação quantitativa das regiões onde a interferência clonal e a deriva genética são dominantes sem nenhuma suposição a priori. Aqui repetimos a mesma análise a fim de elucidarmos um pouco a questão da dinâmica de fixação de mutações benéficas

em populações estruturadas.



**Figura 5.7** Distribuição de probabilidades de tamanhos de população máximos alcançado pelas mutações que foram posteriormente eliminadas. Os parâmetros são  $N = 10000$ ,  $D = 200$ ,  $z = 4$ ,  $e = 0$  e  $m = 0.0001$  (linha tracejada curta),  $m = 0.001$  (linha tracejada) e  $m = 0.01$  (linha cheia).

A Figura 5.6 mostra a distribuição de probabilidades  $P(s)$  de tamanhos máximos de população atingidos por mutações benéficas antes de serem extintas. Nessas simulações consideramos populações compostas por  $N = 10,000$  indivíduos e que estão subdivididas em  $D = 200$  demes. Na Figura temos taxa de migração  $m = 0.0001$ . Em oposição ao que foi observado para populações homogêneas ou espacialmente estruturadas, a distribuição  $P(s)$  é caracterizada pela ocorrência de diversos picos distribuídos sobre um linha reta em uma escala log-log. Constatamos também que estes picos se manifestam em torno dos múltiplos inteiros do tamanho da deme  $N_d = 50$ . Na Figura 5.7 realizamos a mesma análise porém agora aumentamos o tamanho da deme para  $N_d = 100$ . Podemos ver novamente alguns picos na distribuição localizados em múltiplos agora do tamanho da deme  $N_d = 100$ . Ainda na Figura 5.7, observamos também que um aumento na taxa de migração  $m$  reduz a amplitude dos picos, e para  $m$  muito grande os picos não são mais notados. Desde que o tempo que um dado indivíduo leva para mover-se da deme original para uma das demes vizinhas é proporcional a  $\frac{1}{m}$ , isto implica que para baixas taxas de



migração este tempo é muito grande quando comparado com o tempo que uma dada mutação vantajosa gasta para atingir a fixação dentro de uma deme. Para este caso, uma dada mutação benéfica pode escapar da deriva e atingir a fixação na deme onde foi originada ou simplesmente se perder no estágio inicial do seu surgimento. Se a mutação benéfica é bem sucedida, pode mover-se para outra deme vizinha e passa a competir para fixar-se nesta segunda deme. Naturalmente, a probabilidade de sobrevivência na segunda deme depende da composição genética dessa deme. Numa subpopulação bem estabelecida com mutações benéficas de grande valor seletivo, a fixação será bastante improvável. Ademais, a população da deme original pode ser invadida e competir com mutantes mais adaptados de demes vizinhas levando à sua extinção na população. De modo geral, esses dois efeitos explicam a existência do primeiro pico. Por outro lado, quando a formação genética da população da segunda deme é mais fraca do que da primeira deme, sua vantagem seletiva pode conduzir ao domínio da segunda deme, e isto explica o surgimento do segundo pico. Este raciocínio pode ser usado para explicar outros picos bem como o espalhamento da mutação benéfica sobre a população inteira em sua fixação final. Para altas taxas de migração, o tempo que uma dada mutação leva para mover-se de uma deme a outra é muito curto. Neste caso, assim que uma mutação vantajosa surge e começa a crescer de tamanho, ela rapidamente se espalha por toda a rede sem necessariamente ter alcançado fixação na primeira e o processo de fixação é mais rápido. Consequentemente os picos em  $P(s)$  desaparecem.



## Conclusões

Neste trabalho estudamos a dinâmica evolucionária em um modelo de população estruturada introduzido por Campos e Gordo (Campos and Gordo 2006), onde a população é subdividida em várias pequenas subpopulações que trocam indivíduos por meio de migração. Diferentemente do modelo original, aqui admitimos que os indivíduos estão sujeitos à influência da seleção natural. Assumimos também a ocorrência de eventos de extinção. Quando uma extinção ocorre os indivíduos dentro da deme são eliminados. Consideramos duas topologias distintas para as redes migratórias: grafos aleatórios e redes livre de escala.

Constatamos de nossas simulações que para taxa de extinção nula, e também assumindo taxa de mutações deletérias igual a zero, a probabilidade de fixação de uma dada mutação como função da vantagem seletiva dessa mutação é independente da taxa de migração como também da topologia da rede. Além do mais, a probabilidade de fixação é exatamente igual àquela obtida para uma população homogênea. Para taxa de extinção nula e taxa de mutação deletéria maior que zero, uma dependência na taxa de migração surge. Quanto menor a taxa de migração  $m$ , menor é a probabilidade de fixação para um dado valor de  $U_d$ . Isto ocorre pelo fato que embora a probabilidade de fixação seja a mesma quando  $e = 0$  e  $U_d = 0$  para qualquer valor de  $m$ , o tempo de fixação muda, pois este é proporcional a  $1/m$ .

Quando consideramos uma taxa de extinção diferente de zero observa-se uma forte dependência da probabilidade de fixação com a migração, como também uma dependência com a topologia da rede. Essa dependência topológica é mais forte para baixas taxas de migração.

Posteriormente realizamos uma análise dinâmica onde permitimos que mutações benéficas ocorressem a uma taxa fixa  $U_b$ . Como esperado, a taxa de fixação de mutações benéficas,  $k_b$ , é uma função crescente de  $U_b$ . Entretanto, vemos também que  $k_b$  decai quando consideramos menores valores de migração  $m$ . Quando medimos  $k_b$ , consideramos uma janela de tempo para a evolução da população e depois medimos as quantidades de interesse. Assim, o tempo

de difusão da mutação benéfica que é proporcional a  $1/m$  desempenha um papel fundamental quando analisamos a velocidade de adaptação da população. O mesmo cenário é visto para taxas de extinção não-nula.

Uma diferença de performance entre as redes de escala livre e grafos aleatórios aparece quando estudamos  $k_b$  no regime grandes valores de  $U_b$ . Esse regime é particularmente importante, pois para altos valores de  $U_b$  existe uma forte competição entre mutações benéficas que surgem em linhagens distintas da população. Essa competição entre mutações é denominada de interferência clonal. Para pequenos valores de  $U_b$  nenhuma diferença entre as topologias é observada. Portanto a topologia da rede tem um papel importante na dinâmica evolutiva da população por aumentar ou diminuir a intensidade da interferência clonal. A investigação desse intercâmbio entre topologias e dinâmica evolutiva de patógenos é essencial para uma melhor estimativa dos parâmetros evolutivos de populações reais.

# Referências

- Albert, R. and A.-L. Barabási (2002). Statistical mechanics of complex networks. *Rev. Mod. Phys.* 74, 47–.
- Albert, R., H. Jeong, and A.-L. Barabási (1999). Diameter of the world-wide web. *Nature* 401, 130–131.
- Barthélemy, M., A. Barrat, R. Pastor-Satorras, and A. Vespignani (2004). *Phys. Rev. Lett.* 92, 178701.
- Bollobás, B. (1981). Degree sequences of random graphs. *Discrete Mathematics* 33, 1.
- Campos, P., J. Combadão, F. Dionísio, and I. Gordo (2006). Muller’s ratchet in random graphs and scale-free networks. *Phys. Rev. E* 74, 041901.
- Campos, P. and I. Gordo (2006). Pathogen genetic variation in small-world host cont structures. *J. Stat. Mech.*.
- Campos, P. R. A. and V. M. de Oliveira (2004). Mutational effects on the clonal interference phenomenon. *Evolution* 58, 932–937.
- Charlesworth, B., M. Morgan, and D. Charlesworth (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics* 134, 1289–1303.
- Connway, S. and et al. (2000). A principal target of human immunity to malaria identified by molecular population genetics and immunological analyses. *Nature Medicine* 6, 689–692.
- Cuevas, J. M., S. F. Elena, and A. Moya (2002). Molecular basis of adaptive convergence in experimental populations of rna viruses. *Genetics* 162, 533–542.
- de Oliveira, V. M. and P. R. A. Campos (2004). Dynamics of fixation of advantageous mutations. *Physica A* 337, 546–554.

- Dobson, A. (2003). Ecology: metalifes! *Science* 301, 1488–1490.
- Erdos, P. and A. Rényi (1959). On random graphs. *Publ. Math. Debrecen* 6, 290–297.
- Falush, D. and et al. (2003). Traces of human migrations in helicobacter pylori populations. *Science* 299, 1582–1585.
- Gerrish, P. J. and R. E. Lenski (1998). The fate of competing beneficial mutations in an asexual population. *Genetica* 102, 127–144.
- Gillespie, J. H. (1991). *The Causes of Molecular Evolution*. Oxford University Press.
- Gordo, I. and P. R. A. Campos (2006). Adaptive evolution in a spatially structured asexual population. *Genetica* 127, 217–229.
- Grenfell, B., O. Pybus, J. Gog, J. Wood, J. Daly, J. Mumford, and E. Holmes (2004). Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* 303, 327–331.
- Haldane, J. B. S. (1927). A mathematical theory of natural and artificial selection. part v: Selection and mutation. *Proc. Camb. Phil. Soc.* 26, 220–230.
- Harris, T. E. (1963). *The Theory of Branching Processes*. Springer.
- Hill, W. G. and A. Robertson (1966). The effect of linkage on the limits to artificial selection. *Genet. Res.* 8, 269–294.
- Keeling, M.J. and Eames, K. (2005). Networks and epidemic models. *J R Soc Interface* 2, 295–307.
- Kimura, M. (1953). The stepping stone model of population. *Annu. Rept. Natl. Inst. Genet. Japn* 3, 62–63.
- Levins, R. (1969). Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the Entomological Society of America* 15, 237–240.
- Lloyd, A. and M. R.M. (2001). How viruses spread among computers and people. *Science* 292, 1316–1317.

- Maruyama, T. (1970a). On the probability of fixation of mutant genes in subdivided populations. *Genetical Research* 15, 221–225.
- Maruyama, T. (1970b). On the probability of fixation of mutant genes in subdivided populations. *Genetical Research* 15, 221–225.
- McDonald, B. and Linde, C (2003). Pathogen population genetics, evolutionary potential, and durable resistance. *Annu. Rev. Phytopathol.* 40, 349–379.
- Milgram, S. (1967). The small world problem. *Psychol. Today.* 2, 60.
- Miralles, R., P. J. Gerrish, A. Moya, and S. F. Elena (1999). Clonal interference and the evolution of RNA viruses. *Science* 285, 1745–1747.
- Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM* 45, 167–256.
- Ohta, T. (1992). The nearly neutral theory of molecular evolution. *Annual Review of Ecology and Systematics* 23, 263–286.
- Orr, H. A. (2000). The rate of adaptation in asexuals. *Genetics* 155, 961–968.
- Orr, H. A. (2003). The distribution of fitness effects among beneficial mutations. *Genetics* 163, 1519–1526.
- Pannell, J. and B. Charlesworth (2000). Effects of metapopulation processes on measures of genetic diversity. *Phil. Trans. R. Soc. Lond. B* 355, 1851–1864.
- Pastor-Satorras, R. and A. Vespignani (2001). Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* 86, 3200–3203.
- Rosas, A., I. Gordo, and P. R. A. Campos (2005). Scaling, genetic drift and clonal interference in the extinction pattern of asexual populations. *Phys. Rev. E* 72, 012901.
- Sella, G. and A. Hirsh (2005). The application of statistical physics to evolutionary biology. *Proc. Natl. Acad. Sci. USA* 102, 9541–9546.
- Simon, D. and B. Derrida (2006). Evolution of the most recent common ancestor of a population with no selection. *J. Stat. Mech.*

- Watterson, G. A. (1975). Number of segregating sites in genetic models without recombination. *Theor. Popul. Biol.* 10, 256–276.
- Watts, D. and S. Strogatz (1998). Collective dynamics of small world networks. *Nature* 393, 440–442.
- Wright, S. (1931a). Evolution in mendelian populations. *Genetics* 16, 97–159.
- Wright, S. (1931b). Evolution in mendelian populations. *Genetics* 16, 97–159.



