

JOSÉ ALVINO DE LIMA FILHO

**ESTUDO DAS PROPRIEDADES ESTATÍSTICAS DE
CAMINHADAS ADAPTATIVAS EM RELEVOS DE
ADAPTAÇÃO CORRELACIONADOS**

RECIFE-PE - JULHO/2011



UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA

**ESTUDO DAS PROPRIEDADES ESTATÍSTICAS DE
CAMINHADAS ADAPTATIVAS EM RELEVOS DE
ADAPTAÇÃO CORRELACIONADOS**

Dissertação apresentada ao Programa de Pós-Graduação em Biometria e Estatística Aplicada como exigência parcial à obtenção do título de Mestre.

Área de Concentração: Modelagem Estatística e Computacional

Orientador: Prof^a. Dra. Viviane Moraes de Oliveira

RECIFE-PE - JULHO/2011.

Ficha Catalográfica

L732e Lima Filho, José Alvino de
Estudo das propriedades estatísticas de caminhadas adaptativas em relevos de adaptação correlacionados / José Alvino de Lima Filho. -- Recife, 2011.
66 f. : il.

Orientador (a): Viviane Moraes de Oliveira.
Dissertação (Mestrado em Biometria e Estatística Aplicada) – Universidade Federal Rural de Pernambuco, Departamento de Estatística e Informática, Recife, 2011.
Referências.

1. Estatística computacional 2. Caminhadas adaptativas
3. Relevos de fitness 4. Biologia evolucionária I. Oliveira, Viviane Moraes de, Orientadora II. Título

CDD 310

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA

**ESTUDO DAS PROPRIEDADES ESTATÍSTICAS DE CAMINHADAS ADAPTATIVAS EM
RELEVOS DE ADAPTAÇÃO CORRELACIONADOS**

José Alvino de Lima Filho

Dissertação julgada adequada para obtenção do título de mestre em Biometria e Estatística Aplicada, defendida e aprovada por unanimidade em 05/07/2011 pela Comissão Examinadora.

Orientador:

Prof^a. Dra. Viviane Moraes de Oliveira
Universidade Federal Rural de Pernambuco

Banca Examinadora:

Prof. Dr. Borko Stosic
Universidade Federal Rural de Pernambuco

Prof^a. Dra. Tatijana Stosic
Universidade Federal Rural de Pernambuco

Prof. Dr. Pedro Hugo de Figueirêdo
Universidade Federal Rural de Pernambuco

Dedico a todas as pessoas que fazem parte de minha vida, direta ou indiretamente, inclusive aquelas que nos permitem melhor exercitar a paciência e a tolerância. Em especial, dedico à minha filha Maria Luíza, pela compreensão, pela paciência e pela admiração com que sempre me presenciava estudando.

Agradecimentos

Inicialmente gostaria de agradecer a Deus, o Divino Criador, pela minha vida, a primeira de todas as oportunidades, e pela saúde física, mental e espiritual que me permitem prosseguir. Agradeço também a Ele por nunca ter me concedido a genialidade, de modo que a cada passo percebo mais claramente a necessidade do esforço próprio em busca da autossuperação.

Agradeço muito fortemente aos meus avós paternos, Emilio e Marié, por terem ido até o fim na sua crença de que os estudos permitem a mobilidade social, acreditando no empenho e lealdade de meu pai, José Alvino de Lima, que veio à cidade grande a trabalho, e em meros dois anos pôde trazer para perto de si seu pai, sua mãe e seus onze irmãos, atitude que veio a definir nossas vidas até hoje.

Agradeço ainda ao meu avô materno, Antônio de Batista Leite, por ter me transmitido pelo exemplo o amor à poesia, à leitura; Agradeço à minha avó materna, Dona Ritinha, por ter me deixado o exemplo de mansidão e doçura, que permanecem em minhas lembranças da sua forma de viver e conviver.

Agradeço muito a meus pais, José Alvino de Lima e Maria Yleda Leite Lima, por todo o suporte que me deram, cuidando e zelando pela minha integridade física, emocional, espiritual. Aquilo que não consegui corresponder foi por minhas próprias dificuldades, pois todas as condições me foram dadas por esse heróis de minha vida.

Às minhas queridas irmãs, Maria Rita, Tininha e Mônica, por tudo o que elas fazem por mim. Elas bem sabem, pois não me canso de agradecer a essas mulheres de minha vida. E ao meu querido sobrinho, Klebinho, pela constante alegria sempre que me vê.

Muito obrigado a minha filha Maria Luíza, por ter iluminado e oxigenado minha existência de forma exponencial. A ti agradeço, minha filha, por tua companhia. Também agradeço à sua mãe, Catharine, pelo suporte e o incentivo nunca negados.

Ao amigo Edgar Gemir, por ser para mim um irmão mais velho, me escutando e me orientando sempre. Um amigo com quem eu posso compartilhar minhas dificuldades e, principalmente, um dos poucos com quem posso compartilhar minhas alegrias.

Meu muito obrigado a todos os amigos, pelos incentivos, pela contribuição que tiveram na minha formação ética, moral, espiritual, científica, política. Mesmo sabendo que lembrarei de algum outro após a impressão dessas páginas, quero deixar aqui meu agra-

decimento a amigos como Emilio Melo, Manoel Alvino Filho, Érico Andrade, Ricardo Bernardo, Leonardo Moraes da Costa, Edgar Gemir, Verônica Pimentel, José Menezes, Arthur Galamba, Leonardo Portela, Luis Portela, Luciano Portela, Marcos Fábio e tantos outros.

Agradeço a Cristiane Campelo, pela paciência e pelo carinho com que corrigiu o texto final e por ter me escutado quando precisei exercitar o que iria dizer durante a defesa da dissertação.

Meu agradecimento ao Instituto Capibaribe, minha primeira Escola e a Dona Raquel, sua idealizadora e diretora, tão responsáveis pelos meus primeiros passos intelectuais em minha vida. À minha primeira Professora de Matemática, a querida Lêda Telles, que soube nos presentear com sua visão sobre educação matemática já tão à frente, naqueles idos anos 80. E à nossa querida Alderita, hoje na Pátria Espiritual, que com sua exigência e carinho tão bem dosados, nos soube preparar a base em nossa amada língua materna.

Agradeço imensamente ao eterno amigo e mestre, Prof. Ednaldo Ernesto, principal motivador pela escolha de minha profissão, companheiro sempre presente em minha vida, e que me serve sempre como exemplo de educador. Um verdadeiro mestre no ensino da Matemática, com uma capacidade inigualável de nos cativar e conduzir ao maravilhoso mundo da Rainha das Ciências.

Por todo o aprendizado que me foi proporcionado no curso de Licenciatura em Matemática na UFPE, agradeço a todos os meus professores. Especial agradecimento aos professores Antônio Carlos, grande incentivador de que eu continuasse os meus estudos, Paulo Figueiredo, Airton Castro e Manoel Lemos. Agradeço ao professor André Toom, por tanto conhecimento passado com muita humildade e sem falsa modéstia, durante disciplina isolada cursada no mestrado em Estatística da UFPE.

Agradeço muito especialmente a Fátima Bacelar, secretária da graduação de Matemática da UFPE, por todas as incontáveis ajudas, lembretes da data da matrícula, pedidos de cartas de recomendação, tudo isso sempre com o sorriso no rosto.

Bom, essas pessoas antes citadas me foram importantes sempre, independentemente da minha decisão em mais estudar. Mas depois que decidi pelo mestrado nesse Programa, sinto prazer em agradecer ao amigo Cícero Carlos Ramos de Brito, por todas as sugestões e incentivos e a todos os colegas que me engrandeceram com sua companhia e seu conhecimento. Impossível registrar todas as perguntas respondidas com tanta paciência por amigos como Paulo, Rita, Aranildo, Rodrigo, Dani, David, Darlon, Jáder, Mariese, Rosilda, Carlos, Kleber, Patrícia, Elias, Felipe, Sílvio, Cintia e tantos outros. Ao nobre professor Gabriel Rivas, que em sua simplicidade, maior prova de sua grandiosidade, presenteou-me a vida com seus gestos de atenção, de verdadeira preocupação com as minhas dúvidas e questionamentos. Pelos seus bons conselhos, eu agradeço. E ao meu amigo de outras

datas, com quem estudei desde o processo seletivo para esse mestrado, meu nobre amigo Rogério Carlos. Valeu companheiro! Quantas horas de estudo! Vamos em frente!

Quanto aos professores, tive a imensa alegria, pela primeira vez em que fui ao Departamento procurar informações sobre o mestrado, de encontrar o Professor Borko Stosic, sempre atencioso e eficaz. Um grande incentivador e pesquisador admirável, que me servirá como referência na conciliação entre pesquisador e homem de bem. Aos demais professores com quem pude conviver nesse departamento, como Prof^a. Tatijana Stosic, Prof. Cláudio, Prof. Eufrázio e o Prof. Antonio Aleixo.

Agradeço imensamente, e não disponho de palavras para registrar minha gratidão e admiração, à minha orientadora, Prof^a. Viviane Oliveira, pela sua competência, organização, disponibilidade, compreensão e tantas outras virtudes como orientadora pelas quais só posso agradecer. Obrigado pela paciência nas correções, pela gentileza quando eram necessárias as cobranças e por todo o conhecimento que pude adquirir durante esse período de convivência sob sua orientação. Agradeço também ao Prof. Paulo Campos por diversas dicas que foram dadas com relação à programação em C e pelo cumprimento sempre atencioso quando me encontrava.

Zuleide! Quem esquece a inesquecível Zuleide? Estarás sempre em meu coração, querida amiga. Obrigado por tanto carinho, sem qualquer interesse em troca. Muito obrigado. Sei que devo lembrar de agradecer pelos cafezinhos, mas eles ficam frios quando comparados com as tuas palavras de incentivo. Mesmo assim, obrigado pelos cafezinhos. Também pelas bolachinhas, vez ou outra, quando eu comentava que estava com fome.

Agradeço a Marco pela sua disponibilidade em nos ajudar em tudo o que era possível, e até no que parecia não ser possível.

Agradeço, enfim, a todos os que me ajudaram a ser quem eu sou.

“Pensem que todas as maravilhas, objetos de seus estudos, são a obra de muitas gerações, uma obra coletiva que exige de todos um esforço entusiasta e um labor difícil e impreterível. Tudo isto, nas mãos de vocês, se torna uma herança. Vocês a recebem, respeitam-na, aumentam-na e, mais tarde, irão transmiti-la fielmente à sua descendência. Deste modo, somos mortais imortais, porque criamos juntos obras que nos sobrevivem. Se refletirem seriamente sobre isto, encontrarão um sentido para a vida e para o progresso. E o julgamento que fizerem sobre os outros homens e as outras épocas será mais verdadeiro.”

Albert Einstein (Como vejo o mundo)

Resumo

O estudo dos processos que regem a evolução das espécies é um tema que desperta grande interesse e investigações. Neste trabalho, investigamos propriedades estatísticas de caminhadas adaptativas sobre relevos de adaptação correlacionados cujos espaços de sequências possuem topologias diferentes, a saber, grafos aleatórios e redes livres de escala. Utilizamos duas distribuições diferentes para gerar o relevo adaptativo, a distribuição normal e a exponencial bivariada. Também fizemos uso de dois algoritmos evolucionários para a realização das caminhadas, a caminhada padrão e a caminhada gradiente. Observamos que o comprimento médio das caminhadas aumenta com a conectividade da rede quando a caminhada padrão é considerada. Este comportamento não varia quando diferentes topologias e diferentes graus de correlação são considerados. Para a caminhada gradiente em relevos correlacionados, notamos um pico no comprimento médio da caminhada para valores pequenos de conectividade e para grandes valores de conectividade verificamos que o comprimento médio das caminhadas é próximo ao valor estimado para caminhadas gradiente em redes regulares, $L = e - 1$. Por meio do cálculo da diferença relativa entre os valores adaptativos dos mínimos locais e global, verificamos que a caminhada gradiente é mais eficiente do que a caminhada padrão, levando o valor adaptativo a ótimos locais mais próximos do ótimo global.

Palavras-chave: Caminhadas adaptativas, relevos de fitness, biologia evolucionária.

Abstract

The understanding of the processes governing the evolution of species is an issue which raises interest and investigations. In this work, we investigate the statistical properties of adaptive walks on correlated fitness landscapes on sequence spaces of different topologies, namely, random graphs and scale-free networks. We have used two distinct distributions in order to generate the fitness landscape, the normal distribution and the bivariate exponential. We have also used two evolutionary algorithms to perform the adaptive walks, the standard adaptive walk and the gradient adaptive walk. We have observed that the mean walk length is increased with the connectivity of the network in the case which the standard adaptive walk is considered. This behavior does not vary when different topologies and different correlations are considered. When the gradient adaptive walk is performed on correlated fitness landscapes, we can notice a peak for the mean walk length for small values of connectivity and for large values of connectivity we verify that the value of the mean walk length is close to the one estimated for gradient adaptive walks on regular lattices, $L = e - 1$. By calculating the relative difference among the adaptive values of the local minima and the global minimum, we have noticed the gradient adaptive walk is more efficient than the standard adaptive walk, leading the adaptive value to local minima which are closer of the global minimum.

Key-words: Random walks, fitness landscapes, evolutionary biology.

Lista de Figuras

2.1	Hipercubos com dimensões (a)1, (b)2, (c)3, e (d)4.	p. 4
2.2	(a) Cubo representando os genótipos; (b) Cubo representando os genótipos e seus fitness; (c) Caminhadas adaptativas partindo do vértice em destaque (amarelo).	p. 6
2.3	Ilustração de um grafo com $N = 6$ nós e $n = 6$ arestas. O conjunto de vértices é $P = \{1, 2, 3, 4, 5, 6\}$ e $E = \{\{1, 2\}, \{1, 3\}, \{2, 4\}, \{2, 5\}, \{3, 6\}, \{4, 6\}\}$ é o conjunto de arestas.	p. 6
2.4	Em (a),(b) e (c) foram construídos os grafos aleatórios para $p = 0.25, 0.5$ e 0.9 , respectivamente.	p. 8
2.5	Distribuição de frequência de graus de uma rede livre de escala, com 2^{15} nós.	p. 9
2.6	Redes livres de escala construídas por simulação computacional. Em (a) temos uma rede com 20 nós, em que $m = 10$. Em (b) temos uma rede com 40 nós, construído a partir de $m = 20$	p. 10
4.1	Em (a) um grafo aleatório com os aglomerados de tamanho pequeno em destaque. Em (b), o grafo obtido após sua eliminação.	p. 17
4.2	Organograma das simulações	p. 19
4.3	Comprimento médio das caminhadas padrão. Em (a) e (b) temos os resultados com a distribuição de fitness normal, para as redes livres de escala e os grafos aleatórios, respectivamente. Em (c) e (d) temos os resultados com a distribuição de fitness exponencial bivariada, para as redes livres de escala e os grafos aleatórios, respectivamente.	p. 22

- 4.4 Comprimento médio das caminhadas gradiente. Em (a) e (b) temos os resultados com a distribuição de fitness normal, para as redes livres de escala e os grafos aleatórios, respectivamente. Em (c) e (d) temos os resultados com a distribuição de fitness exponencial bivariada, para as redes livres de escala e os grafos aleatórios, respectivamente. p. 23
- 4.5 Densidade média de ótimos locais nas redes. Em (a) e (b) temos os resultados com a distribuição de fitness normal, para as redes livres de escala e os grafos aleatórios, respectivamente. Em (c) e (d) temos os resultados com a distribuição de fitness exponencial bivariada, para as redes livres de escala e os grafos aleatórios, respectivamente. p. 24
- 4.6 Resultados de $\overline{\Delta F_G}$ em função de z , com a distribuição de fitness normal. Em (a) e (b) temos os resultados nas redes livres de escala, para as caminhadas padrão e gradiente, respectivamente. Em (c) e (d) temos os resultados nos grafos aleatórios, para as caminhadas padrão e gradiente, respectivamente. p. 25
- 4.7 Resultados de $\overline{\Delta F_G}$ em função de z , com a distribuição de fitness exponencial bivariada. Em (a) e (b) temos os resultados nas redes livres de escala, para as caminhadas padrão e gradiente, respectivamente. Em (c) e (d) temos os resultados nos grafos aleatórios, para as caminhadas padrão e gradiente, respectivamente. p. 26
- 4.8 Gráficos comparativos dos valores de $\overline{\Delta F_G}$ em função de z , para as caminhadas padrão (quadrado preto) e gradiente (círculo vermelho), com distribuição normal de fitness. Nas linhas, de cima para baixo, temos respectivamente, as caminhadas nas redes livres de escala e nos grafos aleatórios. Nas colunas, da esquerda para a direita, temos, respectivamente, $\lambda = 0, 0.5, 0.9, 0.95$ e 0.99 p. 28
- 4.9 Gráficos comparativos dos valores de $\overline{\Delta F_G}$ em função de z , para as caminhadas padrão (quadrado preto) e gradiente (círculo vermelho), com distribuição exponencial de fitness. Nas linhas, de cima para baixo, temos respectivamente, as caminhadas nas redes livres de escala e nos grafos aleatórios. Nas colunas, da esquerda para a direita, temos, respectivamente, $\theta = 0, 0.5, 0.9, 0.95$ e 0.99 p. 29

- 4.10 Distribuição de frequência dos números de passos das caminhadas padrão nas redes livres de escala, com distribuição normal de fitness. Considerando as linhas de cima para baixo, temos $z = 2, 4, 12, 40$ e 136 . Considerando as colunas da esquerda para a direita, temos $\lambda = 0, 0.5, 0.9, 0.95$ e 0.99 . Em cada gráfico, no eixo horizontal temos o número de passos das caminhadas. p. 30
- 4.11 Distribuição de frequência dos números de passos das caminhadas padrão nos grafos aleatórios, com distribuição normal de fitness. Considerando as linhas de cima para baixo, temos $z = 2, 4, 12, 40$ e 136 . Considerando as colunas da esquerda para a direita, temos $\lambda = 0, 0.5, 0.9, 0.95$ e 0.99 . Em cada gráfico, no eixo horizontal temos o número de passos das caminhadas. p. 31
- 4.12 Distribuição de frequência dos números de passos das caminhadas padrão nas redes livres de escala, com distribuição de fitness exponencial bivariada. Considerando as linhas de cima para baixo, temos $z = 2, 4, 12, 40$ e 136 . Considerando as colunas da esquerda para a direita, temos $\theta = 0, 0.5, 0.9, 0.95$ e 0.99 . Em cada gráfico, no eixo horizontal temos o número de passos das caminhadas. p. 32
- 4.13 Distribuição de frequência dos números de passos das caminhadas padrão nos grafos aleatórios, com distribuição de fitness exponencial bivariada. Considerando as linhas de cima para baixo, temos $z = 2, 4, 12, 40$ e 136 . Considerando as colunas da esquerda para a direita, temos $\theta = 0, 0.5, 0.9, 0.95$ e 0.99 . Em cada gráfico, no eixo horizontal temos o número de passos das caminhadas. p. 33
- 4.14 Distribuição de frequência dos números de passos das caminhadas gradiente nas redes livres de escala, com distribuição normal de fitness. Considerando as linhas de cima para baixo, temos $z = 2, 4, 12, 40$ e 136 . Considerando as colunas da esquerda para a direita, temos $\lambda = 0, 0.5, 0.9, 0.95$ e 0.99 . Em cada gráfico, no eixo horizontal temos o número de passos das caminhadas. p. 34

- 4.15 Distribuição de frequência dos números de passos das caminhadas gradiente nos grafos aleatórios, com distribuição normal de fitness. Considerando as linhas de cima para baixo, temos $z = 2, 4, 12, 40$ e 136 . Considerando as colunas da esquerda para a direita, temos $\lambda = 0, 0.5, 0.9, 0.95$ e 0.99 . Em cada gráfico, no eixo horizontal temos o número de passos das caminhadas. p. 35
- 4.16 Distribuição de frequência dos números de passos das caminhadas gradiente nas redes livres de escala, com distribuição de fitness exponencial bivariada. Considerando as linhas de cima para baixo, temos $z = 2, 4, 12, 40$ e 136 . Considerando as colunas da esquerda para a direita, temos $\theta = 0, 0.5, 0.9, 0.95$ e 0.99 . Em cada gráfico, no eixo horizontal temos o número de passos das caminhadas. p. 36
- 4.17 Distribuição de frequência dos números de passos das caminhadas gradiente nos grafos aleatórios, com distribuição de fitness exponencial bivariada. Considerando as linhas de cima para baixo, temos $z = 2, 4, 12, 40$ e 136 . Considerando as colunas da esquerda para a direita, temos $\theta = 0, 0.5, 0.9, 0.95$ e 0.99 . Em cada gráfico, no eixo horizontal temos o número de passos das caminhadas. p. 37
- 4.18 Rank das caminhadas padrão nas redes livres de escala, com distribuição normal de fitness. Para as figuras (a), (b), (c), (d) e (e) temos $\lambda = 0, 0.5, 0.9, 0.95$ e 0.99 , respectivamente. Em cada gráfico, para cada cor corresponde um valor de z , a saber: $z = 2$ (preto), $z = 4$ (cinza escuro), $z = 6$ (cinza claro), $z = 8$ (vermelho), $z = 14$ (laranja), $z = 40$ (azul claro), $z = 72$ (azul escuro), $z = 104$ (verde claro), $z = 136$ (verde escuro). p. 38
- 4.19 Rank das caminhadas padrão nos grafos aleatórios, com distribuição normal de fitness. Para as figuras (a), (b), (c), (d) e (e) temos $\lambda = 0, 0.5, 0.9, 0.95$ e 0.99 , respectivamente. Em cada gráfico, para cada cor corresponde um valor de z , a saber: $z = 2$ (preto), $z = 4$ (cinza escuro), $z = 6$ (cinza claro), $z = 8$ (vermelho), $z = 14$ (laranja), $z = 40$ (azul claro), $z = 72$ (azul escuro), $z = 104$ (verde claro), $z = 136$ (verde escuro). p. 39

- 4.20 Rank das caminhadas padrão nas redes livres de escala, com distribuição de fitness exponencial bivariada. Para as figuras (a), (b), (c), (d) e (e) temos $\theta = 0, 0.5, 0.9, 0.95$ e 0.99 , respectivamente. Em cada gráfico, para cada cor corresponde um valor de z , a saber: $z = 2$ (preto), $z = 4$ (cinza escuro), $z = 6$ (cinza claro), $z = 8$ (vermelho), $z = 14$ (laranja), $z = 40$ (azul claro), $z = 72$ (azul escuro), $z = 104$ (verde claro), $z = 136$ (verde escuro). p. 40
- 4.21 Rank das caminhadas padrão nos grafos aleatórios, com distribuição de fitness exponencial bivariada. Para as figuras (a), (b), (c), (d) e (e) temos $\theta = 0, 0.5, 0.9, 0.95$ e 0.99 , respectivamente. Em cada gráfico, para cada cor corresponde um valor de z , a saber: $z = 2$ (preto), $z = 4$ (cinza escuro), $z = 6$ (cinza claro), $z = 8$ (vermelho), $z = 14$ (laranja), $z = 40$ (azul claro), $z = 72$ (azul escuro), $z = 104$ (verde claro), $z = 136$ (verde escuro). p. 41
- 4.22 Rank das caminhadas gradiente nas redes livres de escala, com distribuição normal de fitness. Para as figuras (a), (b), (c), (d) e (e) temos $\lambda = 0, 0.5, 0.9, 0.95$ e 0.99 , respectivamente. Em cada gráfico, para cada cor corresponde um valor de z , a saber: $z = 2$ (preto), $z = 4$ (cinza escuro), $z = 6$ (cinza claro), $z = 8$ (vermelho), $z = 14$ (laranja), $z = 40$ (azul claro), $z = 72$ (azul escuro), $z = 104$ (verde claro), $z = 136$ (verde escuro). p. 42
- 4.23 Rank das caminhadas gradiente nos grafos aleatórios, com distribuição normal de fitness. Para as figuras (a), (b), (c), (d) e (e) temos $\lambda = 0, 0.5, 0.9, 0.95$ e 0.99 , respectivamente. Em cada gráfico, para cada cor corresponde um valor de z , a saber: $z = 2$ (preto), $z = 4$ (cinza escuro), $z = 6$ (cinza claro), $z = 8$ (vermelho), $z = 14$ (laranja), $z = 40$ (azul claro), $z = 72$ (azul escuro), $z = 104$ (verde claro), $z = 136$ (verde escuro). p. 43
- 4.24 Rank das caminhadas gradiente nas redes livres de escala, com distribuição de fitness exponencial bivariada. Para as figuras (a), (b), (c), (d) e (e) temos $\theta = 0, 0.5, 0.9, 0.95$ e 0.99 , respectivamente. Em cada gráfico, para cada cor corresponde um valor de z , a saber: $z = 2$ (preto), $z = 4$ (cinza escuro), $z = 6$ (cinza claro), $z = 8$ (vermelho), $z = 14$ (laranja), $z = 40$ (azul claro), $z = 72$ (azul escuro), $z = 104$ (verde claro), $z = 136$ (verde escuro). p. 44

4.25 Rank das caminhadas gradiente nos grafos aleatórios, com distribuição de fitness exponencial bivariada. Para as figuras (a), (b), (c), (d) e (e) temos $\theta = 0, 0.5, 0.9, 0.95$ e 0.99 , respectivamente. Em cada gráfico, para cada cor corresponde um valor de z , a saber: $z = 2$ (preto), $z = 4$ (cinza escuro), $z = 6$ (cinza claro), $z = 8$ (vermelho), $z = 14$ (laranja), $z = 40$ (azul claro), $z = 72$ (azul escuro), $z = 104$ (verde claro), $z = 136$ (verde escuro). p. 45

Sumário

1	INTRODUÇÃO	p. 1
2	REVISÃO DE LITERATURA	p. 3
2.1	Caminhadas adaptativas	p. 4
2.2	Redes Complexas	p. 6
2.2.1	Grafos Aleatórios	p. 7
2.2.2	A rede livre de escala	p. 8
3	RELEVOS NÃO CORRELACIONADOS	p. 11
4	RELEVOS CORRELACIONADOS	p. 15
4.1	A construção das redes	p. 15
4.1.1	Redes livres de escala	p. 15
4.1.2	Grafos aleatórios	p. 16
4.2	A atribuição dos fitness	p. 16
4.2.1	Distribuição normal	p. 17
4.2.2	Distribuição exponencial bivariada	p. 18
4.3	As propriedades estudadas	p. 19
4.4	Resultados das caminhadas	p. 21
4.4.1	Comprimento médio das caminhadas não-nulas	p. 21
4.4.2	Densidade média de ótimos locais	p. 21
4.4.3	Diferença relativa entre os fitness local e global ΔF_G	p. 22
4.4.4	Distribuição de frequências dos comprimentos das caminhadas	p. 26

4.4.5 Rank das caminhadas p.27

Referências p.48

1 INTRODUÇÃO

Apesar dos avanços recentes em biologia experimental (Elena e Lenski, 1997; Sanjuán et al, 2004), ainda não há uma compreensão total dos mecanismos fundamentais que levam à evolução adaptativa. Neste contexto, investigações teóricas podem ser de grande importância para a biologia evolucionária.

A visualização mais simples para evolução adaptativa é devida a Wright (1931), que introduziu a metáfora de uma escalada de montanha para a evolução Darwiniana e também o conceito de relevo de adaptação. Como resultado de um intenso desenvolvimento da Biologia Evolucionária, a partir da década de 1980 os estudos sobre as caminhadas adaptativas em espaços de sequências discretos foram intensificados. Entende-se que a evolução de uma população é melhor ilustrada por uma caminhada adaptativa da população em um espaço de sequências discreto, em que essas sequências representam os genótipos. A cada passo dessa caminhada a população sofre uma mutação de um gene, passando para um genótipo com maior valor adaptativo ou *fitness*. Destaca-se, nessa década de 1980, o trabalho de Kauffman e Levin (1987), o qual introduziu a ideia de relevo de fitness com variação da rugosidade. Nesse trabalho, foram feitos estudos das caminhadas adaptativas, restringindo-se a espaços de sequências representados por hipercubos (redes regulares). Nesse caso, todos os genótipos possuem o mesmo número de vizinhos na rede.

Campos e Moreira (2005) estudaram as propriedades de caminhadas adaptativas sobre um relevo de adaptação descorrelacionado que é estabelecido no espaço de sequências de estrutura complexa. Nessas investigações foram realizadas simulações de caminhadas adaptativas sobre grafos aleatórios (Erdős e Rényi, 1960) e redes livres de escala (Albert e Barabási, 2002; Newman, 2001). Como dissemos, os estudos anteriores consideravam espaços de sequências representados por hipercubos onde cada mutante está conectado ao mesmo número de vizinhos por 1 mutação, ou seja, cada nó possui a mesma conectividade. Entretanto, uma grande quantidade de mutações têm efeitos letais e dessa forma as sequências podem não permanecer viáveis após sofrerem mutações pontuais

(Lenski et al, 1999). Portanto, não é esperado que cada nó tenha o mesmo número de variantes viáveis.

No entanto, relevos de adaptação realísticos possuem uma topologia mais complexa do que o modelo descorrelacionado. Relevos de RNA (Fontana et al, 1989; Fontana et al, 1993) e bactérias (Lenski e Travisano, 1994) apresentam um alto grau de rugosidade com um grande número de ótimos locais, mas não são exatamente descorrelacionados. Em relevos muito rugosos, os valores adaptativos de sequências com pequena distância de Hamming não são fortemente correlacionados, e assim qualquer mutação pode alterar consideravelmente a adaptação dos organismos. Sabe-se que a correlação do relevo de adaptação pode agir sobre a dinâmica de fixação de mutações vantajosas (de Oliveira e Campos, 2005).

Nessa dissertação, fizemos um estudo da evolução adaptativa sobre relevos rugosos. Mais especificamente, investigamos as propriedades estatísticas de caminhadas adaptativas em um relevo de adaptação correlacionado onde o grau de correlação pode ser ajustado variando-se um parâmetro pré-definido (Wilke et al, 2002). Também investigamos se o processo adaptativo é dependente da escolha particular do algoritmo evolucionário.

No capítulo 2, descrevemos as caminhadas adaptativas, os relevos de fitness, bem como as redes livres de escala e grafos aleatórios. Enfatizamos o processo de construção dessas redes complexas utilizadas em nosso trabalho e a diferença entre os dois tipos de algoritmos considerados, a caminhada padrão e a caminhada gradiente.

No capítulo 3, estudamos alguns trabalhos que foram feitos sobre caminhadas adaptativas em relevos de adaptação não correlacionados. Demos ênfase ao trabalho já citado de Campos e Moreira (2005), enfatizando a sua função de motivador do nosso trabalho.

No capítulo 4, detalhamos todo o nosso estudo feito sobre as caminhadas adaptativas em relevos de adaptação correlacionados. Consideramos caminhadas padrão e gradiente, utilizando as redes livres de escala e os grafos aleatórios, tomando para distribuição de fitness as distribuições normal e exponencial bivariada, com diferentes parâmetros de correlação. Nos interessamos em propriedades que dizem respeito ao número médio de passos nas caminhadas, à distribuição dos ótimos locais nas redes, bem como qual o tipo de caminhada mais eficiente.

No capítulo 5, apresentamos nossas conclusões.

2 REVISÃO DE LITERATURA

Segundo Ronald A. Fisher (1930), “a adaptação é caracterizada pelo movimento de uma população para um fenótipo que melhor se adapta ao ambiente atual”. A Teoria da adaptação foi bastante estudada no Século XX, tendo passado por vários momentos de grandes mudanças, onde algumas teorias caíram com o surgimento de outras.

A primeira visão das bases genéticas da adaptação foi pré-mendeliana. Esse ponto de vista que enfatizava a extrema gradualidade da evolução fenotípica, começou com o próprio Charles Darwin (1859), que argumentou que “a seleção natural só pode ocorrer através de pequenas variações sucessivas, nunca dando saltos, sempre avançando em pequenos e lentos passos”. Apesar de Darwin não ter um correto entendimento da natureza da herança genética, ele concluiu que a base hereditária da evolução adaptativa era bastante refinada, de modo que os organismos se adaptariam aos seus ambientes através de muitos ajustes pequenos. Essa visão micromutacional da adaptação, a qual estabeleceu as bases para a fundação da Escola Biométrica Britânica de Evolução (British Biometric School of Evolution), liderada por Karl Pearson e Walter Weldon, sofreu fortes oposições da escola mendeliana, mas se consolidou entre os evolucionistas por volta de 1930. Essa “vitória” reflete os esforços de Fisher, um dos fundadores da genética populacional e incansável defensor do gradualismo darwinista. Fisher fundiu com sucesso o micromutacionismo e o mendelismo, produzindo uma estrutura matemática conhecida como modelo infinitesimal.

Embora o modelo de Fisher levasse em consideração a natureza mendeliana da mutação, ele não refletia a base molecular da herança, ou seja, não considerava o fato de que o DNA é uma sequência linear de nucleotídeos que, entre outras coisas, codifica uma sequência linear de aminoácidos. Foi quando, no início dos anos 60, os teóricos começaram a desenvolver modelos de adaptação baseados em sequências. John Maynard Smith (1962) enfatizou que a real adaptação ocorre num espaço de sequências que, diferentemente do modelo infinitesimal, é discreto. É nesse período que surgem as caminhadas adaptativas em espaços de sequências discretos, ideias que ficaram quase totalmente ignoradas por duas décadas.

Na década de 1980, o estudo das caminhadas adaptativas foi finalmente retomado com o importante trabalho de Kauffman e Levin (1987), dentre outros que o sucederam, como Orr (2003), Orr (2005), Kauffman (1993), enfatizando, a partir daí, a ideia de relevo de fitness, tema que será estudado nesse nosso trabalho.

2.1 Caminhadas adaptativas

É de comum acordo que as espécies evoluem devido à sua capacidade de adaptação ao meio em que vivem. Nesse sentido, Wright (1931) criou a metáfora da “escalada” da evolução darwiniana. Em outras palavras: a evolução de uma população é melhor ilustrada por uma caminhada adaptativa da população em um espaço de sequências discreto, em que essas sequências representam os genótipos. A cada passo dessa caminhada, a população sofre uma mutação de um gene, passando para um genótipo com maior valor adaptativo ou *fitness*.

Nos trabalhos de Kauffman e Levin (1987), Kauffman (1993) e Orr (2003) foram feitos estudos das caminhadas adaptativas, restringindo-se a espaços de sequências representados por hipercubos (redes regulares). Nesse caso, todos os genótipos possuem o mesmo número de vizinhos na rede.

Para entendermos melhor as caminhadas adaptativas em um espaço de sequências, consideremos um hipercubo de dimensão 3 (um cubo). Cada genótipo é representado por uma sequência binária localizada sobre um vértice do hipercubo. No total temos 2^3 vértices cujas sequências são: (1, 1, 1), (1, 1, 0), (1, 0, 1), (1, 0, 0), (0, 1, 1), (0, 1, 0), (0, 0, 1) e (0, 0, 0). Cada termo da sequência representa um gene, e a título de simplificação os genes só assumem os valores 0 e 1. Sequências que diferem de apenas um gene ficam associadas a vértices vizinhos (que estão interligados por uma aresta). A Figura 2.1 ilustra hipercubos com dimensões 1, 2, 3, e 4.

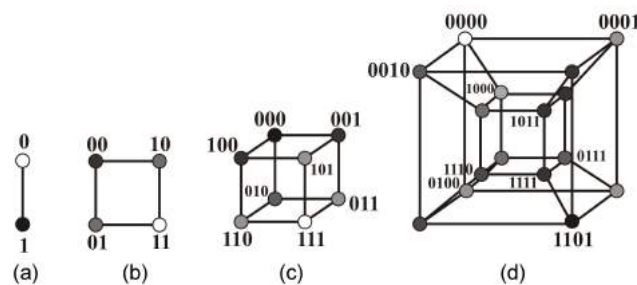


Figura 2.1: Hipercubos com dimensões (a)1, (b)2, (c)3, e (d)4.

Na Figura 2.2(a) temos um cubo com seus vértices num plano (para a realização das caminhadas são importantes apenas como os vértices estão conectados, não importando como se dispõem no espaço euclidiano). A cada vértice (ou genótipo) está associado um valor adaptativo, ou *fitness*. A Figura 2.2(b) ilustra o cubo com um *fitness* associado a cada genótipo, criando a ideia de um relevo adaptativo.

A partir de um vértice desse cubo, um passo pode ser dado para outro vértice ao qual esteja associado um *fitness* maior do que o *fitness* associado ao vértice de origem. Dessa forma são realizadas as caminhadas, até que se atinja um ótimo local, ou seja, um vértice que não possui vizinhos com valores adaptativos maiores do que o seu. A Figura 2.2(c) ilustra possíveis caminhadas partindo do vértice em destaque.

Nesse estudo serão considerados dois tipos de caminhadas adaptativas:

1. *Caminhada padrão*: sorteamos um primeiro nó e verificamos quais são os seus vizinhos mais próximos que possuem valor adaptativo maior que o seu; entre esses (se houver mais de um) sorteamos um, todos com igual probabilidade; um passo é dado para esse nó e repete-se o processo, até que um ótimo local seja encontrado.
2. *Caminhada gradiente*: sorteamos um primeiro nó e verificamos seus vizinhos mais próximos que possuem valor adaptativo maior que o seu; entre esses (se houver mais de um) escolhemos o que possui maior valor adaptativo; um passo é dado para esse nó e repete-se o processo, até que um ótimo local seja encontrado.

Lensky, Ofria, Collier e Adami (1999) e Elena e Lensky (1997) verificaram que nem todas as sequências no espaço de genótipos são viáveis, já que muitas vezes um considerável percentual das mutações são letais. Nesse sentido, Campos e Moreira (2005) realizaram caminhadas adaptativas utilizando redes com uma topologia mais complexa: **as redes livres de escala** e os **grafos aleatórios**. Nessas redes, os números de conexões por nó são diferentes, tornando o modelo mais realístico, já que há mutações não viáveis. Nele foram considerados relevos de adaptação **não-correlacionados**, ou seja, em que os valores dos *fitness* atribuídos aos vértices não têm correlação com os valores atribuídos aos vértices vizinhos. Os valores dos *fitness* foram obtidos por meio de uma distribuição uniforme $(0, 1]$, e foram feitos estudos sobre propriedades estatísticas dessas caminhadas adaptativas. No próximo capítulo descreveremos com mais detalhes os resultados obtidos por Campos e Moreira (2005).

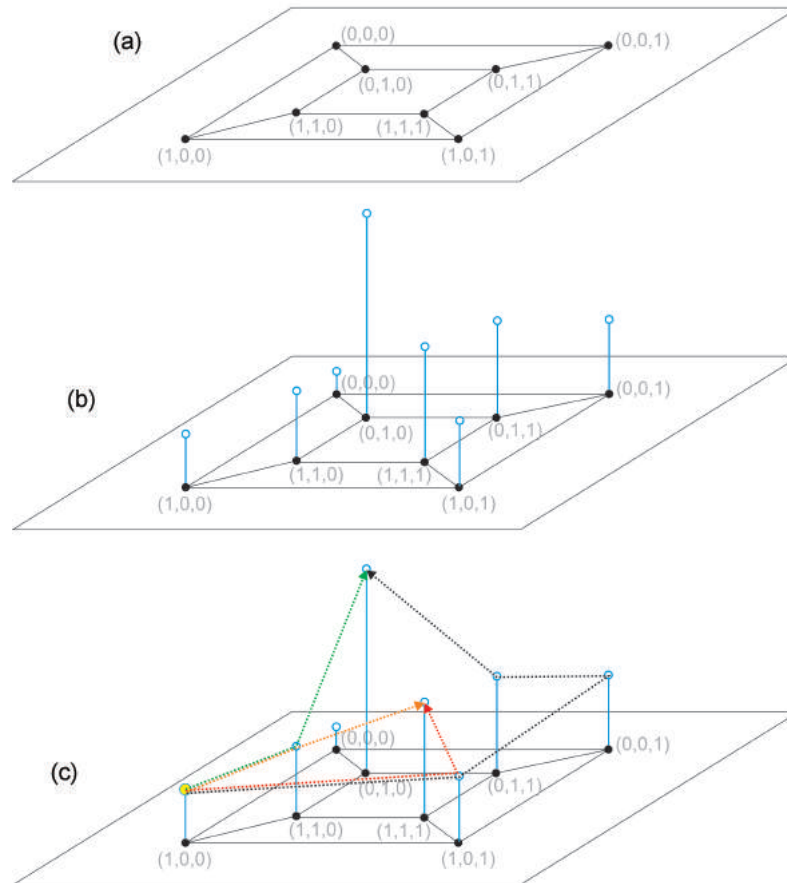


Figura 2.2: (a) Cubo representando os genótipos; (b) Cubo representando os genótipos e seus fitness; (c) Caminhadas adaptativas partindo do vértice em destaque (amarelo).

2.2 Redes Complexas

Basicamente, um grafo é um conjunto de pontos (denominados nós ou vértices), em que alguns desses são conectados por arestas (segmentos de reta). Mais precisamente um grafo é um par de conjuntos $G = \{P, E\}$, em que P é um conjunto de N nós P_1, P_2, \dots, P_N e E é um conjunto de arestas que conectam dois elementos de P . A Figura 2.3 ilustra um grafo com 6 nós e 6 arestas.

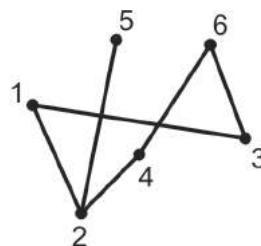


Figura 2.3: Ilustração de um grafo com $N = 6$ nós e $n = 6$ arestas. O conjunto de vértices é $P = \{1, 2, 3, 4, 5, 6\}$ e $E = \{\{1, 2\}, \{1, 3\}, \{2, 4\}, \{2, 5\}, \{3, 6\}, \{4, 6\}\}$ é o conjunto de arestas.

Descreveremos a seguir os tipos de grafos que utilizamos em nosso estudo.

2.2.1 Grafos Aleatórios

Grafo aleatório é todo grafo cujos vértices são conectados aleatoriamente. A teoria dos grafos aleatórios foi introduzida pelos matemáticos húngaros Paul Erdős e Alfréd Rényi (1960), quando Erdős percebeu que os métodos probabilísticos seriam úteis ao estudo desse tipo de grafo (Albert e Barabási, 2002; Newman, 2001).

Para entender os grafos aleatórios, consideremos um grafo com N vértices. Como existem ao todo $N(N-1)/2$ possíveis arestas, o número de possíveis grafos com n arestas será dado por $C_{N(N-1)/2}^n$, em que C_n^p representa o número de combinações de n elementos tomados p a p . Podemos estudar, por exemplo, a probabilidade de ser formado um grafo com determinada propriedade especial ou procurar maximizar a probabilidade de ocorrência dessa propriedade. Mais especificamente, a Teoria dos grafos aleatórios estuda as propriedades do espaço probabilístico associado a grafos com N vértices, quando $N \rightarrow \infty$. Nesse sentido dizemos que *quase todo grafo* possui a propriedade Q , quando a probabilidade dele possuir tal propriedade tende a 1, quando $N \rightarrow \infty$.

Por definição, o **grau** de um vértice é o número de arestas que partem desse vértice. Podemos estudar, por exemplo, a probabilidade de que o grau médio de um grafo assumira um valor r . Outra característica importante dos grafos aleatórios é a existência de **subgrafos**. Dizemos que um grafo G_1 é um subgrafo do grafo G quando todos os vértices de G_1 são vértices de G e todas as arestas de G_1 são também arestas de G . Nesse caso escrevemos: $G_1 \subset G$. Podemos estudar, inclusive, a probabilidade de ocorrência de certos subgrafos específicos.

Uma forma interessante de se pensar a construção de um grafo é o processo comumente chamado de **evolução**: partimos inicialmente dos N vértices do grafo e estabelecemos que a probabilidade de um vértice se conectar a outro seja p . A cada passo de tempo um vértice será conectado a outro com probabilidade p , até que sejam tomados todos os possíveis pares de vértices. O número total de arestas será uma variável aleatória n com esperança $E(n) = pN(N-1)/2$. Ao final desse processo, cada grafo específico terá uma probabilidade de ocorrência, por exemplo:

- A probabilidade de ocorrência de um grafo específico G_0 com n arestas será

$$P(G_0) = p^n \cdot (1-p)^{\frac{N(N-1)}{2} - n}.$$

- A probabilidade de ocorrência de um grafo qualquer G_0 com n arestas será

$$P(G_0) = C_{\frac{N(N-1)}{2}}^n \cdot p^n \cdot (1-p)^{\frac{N(N-1)}{2}-n}.$$

Na Figura 2.4, mostramos uma ilustração de três grafos aleatórios com 10 vértices, construídos a partir de diferentes valores da probabilidade de conexão p .

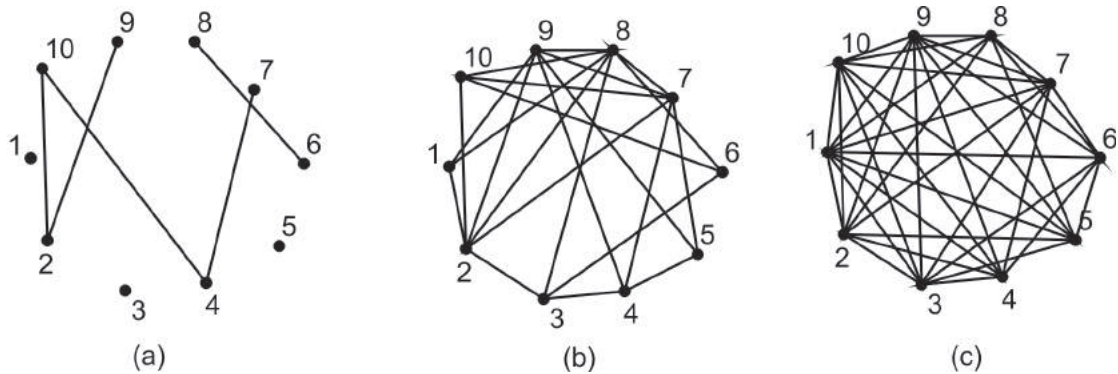


Figura 2.4: Em (a),(b) e (c) foram construídos os grafos aleatórios para $p = 0.25, 0.5$ e 0.9 , respectivamente.

Variando o valor da probabilidade p e o número de nós N , podemos obter grafos com conectividades médias diferentes.

2.2.2 A rede livre de escala

Em diversas redes no mundo real, a quantidade de elementos que compõem a rede pode variar ao longo do tempo. Redes de relacionamento na internet, redes de pessoas que são contaminadas por outras, redes de relacionamentos sexuais, dentre muitos outros exemplos, demonstram a necessidade de se conceber um modelo de rede em que o número de nós possa variar. Uma outra característica presente em redes reais é o fato de que as ligações entre os vértices podem ter probabilidades diferentes, ou seja, de que existem ligações preferenciais (Albert e Barabási, 2002; Newman, 2001).

Definiremos agora o algoritmo do modelo de rede livre de escala, o qual contempla essas duas características anteriormente citadas:

1. **Crescimento:** Começamos com um pequeno número m_0 de vértices, e a cada passo de tempo adicionamos um novo vértice com m arestas, que serão conectadas a m diferentes vértices já presentes no sistema.

2. **Ligações preferenciais:** Ao escolher os vértices aos quais o novo vértice n será conectado, assumimos que a probabilidade Π dele ser conectado ao vértice i será proporcional ao grau k_i desse vértice, ou seja,

$$\Pi_{k_i} = \frac{k_i}{\sum_{j=1}^{n-1} k_j}.$$

Após t passos de tempo, este algoritmo resulta em uma rede com $N = t + m_0$ vértices e mt arestas. Simulações numéricas indicam que esse tipo de rede evolui de modo que a distribuição de graus segue uma lei de potência, ou seja, a probabilidade de um vértice i estar conectado a k arestas é da forma:

$$P(X_i = k) = ck^{-\gamma},$$

em que $\gamma = 3$, quando o número de vértices N tende a infinito.

A seguir, mostramos a Figura 2.5 com a distribuição de frequência de graus de uma rede livre de escala, com 2^{15} nós, construída através de simulação computacional, utilizando o procedimento citado anteriormente. Na simulação, calculamos a frequência relativa $f(k)$ de cada possível grau k , para cada valor de k . Em outras palavras, $f(4) = 0.35$ significa que 35% dos nós possuem grau 4, ou seja, existem 4 arestas conectadas a cada um deles. Fizemos o ajuste linear e encontramos para c e γ os valores $c = 4,002$ e $\gamma = 2,9139$.

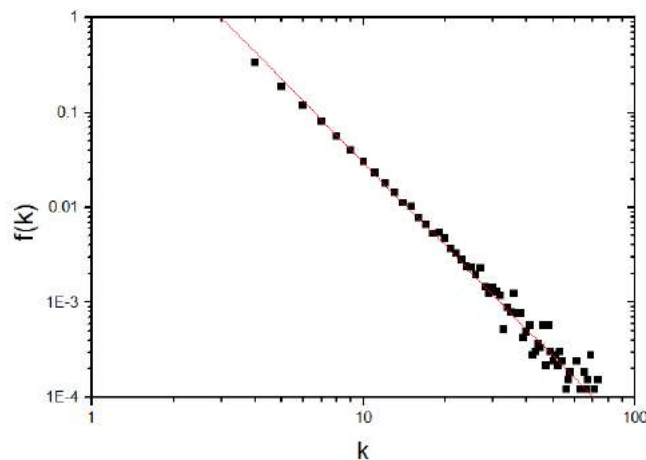


Figura 2.5: Distribuição de frequência de graus de uma rede livre de escala, com 2^{15} nós.

Na figura 2.6, mostramos duas redes livres de escala, construídas a partir de diferentes valores de m e m_0 .

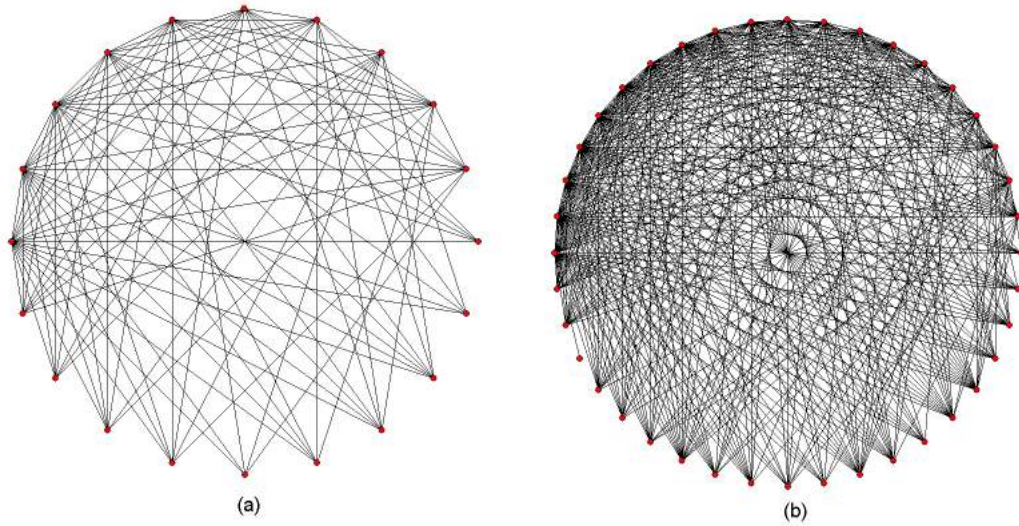


Figura 2.6: Redes livres de escala construídas por simulação computacional. Em (a) temos uma rede com 20 nós, em que $m = 10$. Em (b) temos uma rede com 40 nós, construído a partir de $m = 20$.

3 RELEVOS NÃO CORRELACIONADOS

Kauffman e Levin (1987) investiram numa Teoria Geral sobre as caminhadas adaptativas, inserindo o conceito de espaço de sequências e a ideia de fitness associado a cada entidade. Nesse trabalho, desenvolveram uma teoria geral para o número de ótimos locais, para o comprimento médio das caminhadas e o número de ótimos locais alternativos acessíveis a um dado nó inicial. Foram consideradas redes regulares (hipercubos) e os relevos foram não-correlacionados.

Campos e Moreira (2005) fizeram um estudo sobre algumas propriedades das caminhadas adaptativas em relevos de adaptação não-correlacionados. Nesse trabalho foram utilizadas redes regulares, redes livres de escala e grafos aleatórios, com um número fixo de nós N e várias conectividades médias z . O número de vizinhos de um nó i é denotado por k_i e, no sentido de definir um relevo adaptativo, foi atribuído a cada nó i um fitness $f_i \in (0, 1]$ através da distribuição uniforme. O conjunto de valores $f_i, i = 1, \dots, N$ caracteriza o relevo de fitness.

Três propriedades estatísticas das caminhadas foram estudadas para cada topologia considerada:

- A densidade média de ótimos locais nas redes (η), ou seja, a razão entre o número de ótimos locais e o número total de nós. Considera-se a média entre os valores de η para cada rede.
- O comprimento médio das caminhadas (\bar{L}), ou seja, o número médio de passos da caminhada até um ótimo local.
- A diferença relativa média (ΔF_G) entre os fitness dos ótimos locais e o fitness do ótimo global, ou seja, foi calculada a média entre os valores de $\Delta F_G = \frac{F_{Global} - F_{Local}}{F_{Global}}$ para cada rede. Nesse artigo, como os fitness foram atribuídos através da dis-

tribuição uniforme $(0, 1]$, o valor esperado do fitness do ótimo local é igual a 1, portanto tal razão ficou reduzida à diferença $\Delta F_G = F_{Global} - F_{Local}$.

Como os fitness f_i são uniformemente distribuídos no intervalo $(0, 1]$, a probabilidade de que um dado nó i tenha fitness maior do que um vizinho seu é f_i . Desse modo, se esse nó i possui k_i vizinhos, a probabilidade de que ele seja um ótimo local é $f_i^{k_i}$. Sendo $P(k)$ a distribuição de probabilidade das conectividades, ou seja, a probabilidade de um nó possuir exatamente k vizinhos, foi calculada a densidade média de ótimos locais como

$$\eta = \sum_k \int_0^1 x^k P(k) dx, \quad (3.1)$$

de modo que para cada tipo de topologia considerada teremos uma distribuição $P(k)$ diferente.

Com relação às redes regulares (hipercubos) de dimensão D , todos os nós possuem D vizinhos, portanto $P(k) = \begin{cases} 1 & \text{se } k = D \\ 0 & \text{se } k \neq D \end{cases}$. Nesse caso, a equação (3.1) fica reduzida a

$$\eta = \int_0^1 x^D \cdot 1 \cdot dx = \frac{1}{D+1}. \quad (3.2)$$

Ficou concluído, portanto, que $\eta = 1/(D+1)$ e que para valores grandes de D , $\eta \approx 1/D$.

O comprimento médio das caminhadas gradiente (\bar{L}) nas redes regulares foi estimado por Orr (2003) no valor de $\bar{L} = e - 1 \approx 1,72$. Ele demonstrou que a probabilidade de uma população dar exatamente n passos é dada por

$$P_n = \frac{n}{(n+1)!}, \quad (3.3)$$

e desse modo

$$\bar{L} = \sum_{n=1}^{\infty} n P(n) = \sum_{n=1}^{\infty} n \frac{n}{(n+1)!} = \sum_{n=1}^{\infty} \frac{n^2}{(n+1)!} \quad (3.4)$$

Calculando o valor do somatório em 3.4, temos:

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{n^2}{(n+1)!} &= \sum_{n=1}^{\infty} \frac{n^2 + n - n - 1 + 1}{(n+1)!} = \sum_{n=1}^{\infty} \frac{n(n+1) - (n+1) + 1}{(n+1)!} = \\
&= \sum_{n=1}^{\infty} \frac{n(n+1)}{(n+1)!} - \sum_{n=1}^{\infty} \frac{n+1}{(n+1)!} + \sum_{n=1}^{\infty} \frac{1}{(n+1)!} \\
&= \sum_{n=1}^{\infty} \frac{1}{(n-1)!} - \sum_{n=1}^{\infty} \frac{1}{n!} + \sum_{n=1}^{\infty} \frac{1}{(n+1)!} = \sum_{n=0}^{\infty} \frac{1}{n!} - \sum_{n=1}^{\infty} \frac{1}{n!} + \sum_{n=2}^{\infty} \frac{1}{n!} \\
&= \sum_{n=0}^{\infty} \frac{1}{n!} - \left(\sum_{n=0}^{\infty} \frac{1}{n!} - \frac{1}{0!} \right) + \left(\sum_{n=0}^{\infty} \frac{1}{n!} - \frac{1}{0!} - \frac{1}{1!} \right)
\end{aligned}$$

e como $\sum_{n=0}^{\infty} \frac{1}{n!} = e$, temos finalmente $\bar{L} = e - (e - 1) + (e - 2) = e - 1 \approx 1,72$.

Conforme vimos no capítulo anterior, nos grafos aleatórios os vértices são conectados entre si com a mesma probabilidade p . Este procedimento produz uma rede estatisticamente homogênea, em que a distribuição das conectividades é dada pela distribuição de Poisson, ou seja,

$$P(k) = \frac{e^{-z} z^k}{k!}, \quad (3.5)$$

em que $z = p(N - 1)$ é a conectividade média, ou seja, o número médio de conexões por nó.

Substituindo a Eq. 3.5 na Eq. 3.1, obtemos para a densidade de ótimos locais nos grafos aleatórios:

$$\begin{aligned}
\eta &= \sum_k \int_0^1 x^k \frac{e^{-z} z^k}{k!} dx = \int_0^1 \sum_k \frac{(xz)^k}{k!} e^{-z} dx \\
\eta &= e^{-z} \int_0^1 \sum_k \frac{(xz)^k}{k!} dx = e^{-z} \int_0^1 e^{xz} dx = e^{-z} \left(\frac{e^z}{z} - \frac{1}{z} \right)
\end{aligned}$$

Daí, finalmente:

$$\eta = \frac{1}{z}(1 - e^{-z}) \quad (3.6)$$

Podemos perceber que a partir de valores intermediários de z , a densidade $\eta \approx 1/z$.

Com relação ao comprimento médio \bar{L} , não foi obtido o valor exato pois não temos a mesma conectividade para todos os nós. Nesse caso, foi possível subestimar \bar{L} , assumindo que, se um nó possui um vizinho que é ótimo local, então um passo é dado para ele. De fato, isso nem sempre ocorre, pois o vizinho de maior fitness pode não ser um ótimo local. Foram consideradas apenas as caminhadas que não partem de um ótimo local. A partir dessas suposições, Campos e Moreira obtiveram que $\bar{L} \geq \frac{e}{e-1}$, para valores grandes de z .

4 RELEVOS CORRELACIONADOS

Neste capítulo, apresentamos um estudo das propriedades estatísticas de caminhadas adaptativas sobre relevos de adaptação **correlacionados** que são estabelecidos em espaços de sequências de estrutura complexa. Consideraremos espaços de sequências com duas topologias diferentes:

1. Redes livres de escala;
2. Grafos aleatórios.

4.1 A construção das redes

Foram construídas redes livres de escala e grafos aleatórios com conectividades médias diferentes. A conectividade média (z) é o número médio de conexões por nó. Consideramos os casos $z = 2, 4, 6, 8, 10, 12, 14, 40, 72, 104$ e 136 . Para cada valor de z foram construídas 1000 redes.

4.1.1 Redes livres de escala

Conforme já vimos na seção 2.2.2, cada rede livre de escala é construída partindo de um número N_0 de nós e a cada passo é acrescentado mais um nó que deverá ser conectado a m nós já existentes, dando preferência aos nós com mais conexões, ou seja, é feito um sorteio de m nós entre os nós já existentes, com probabilidades proporcionais aos graus de cada nó.

Construímos redes para $m = 1, 2, 3, 4, 5, 6, 7, 20, 36, 52$ e 68 , obtendo redes com conectividades médias respectivamente iguais a $z = 2, 4, 6, 8, 10, 12, 14, 40, 72, 104$ e 136 . Para cada valor de m tomamos $N_0 = m$ (número inicial de nós) e no primeiro passo acrescentamos um nó, conectando-o a todos os m nós iniciais. A partir do segundo passo,

cada nó acrescentado foi conectado a m nós através das ligações preferenciais, conforme explicado anteriormente. Repetimos esse processo até que a rede possua $2^{15} = 32768$ nós.

4.1.2 Grafos aleatórios

A construção dos grafos aleatórios foi feita baseada no processo de **evolução** citado na seção 2.2.1: partimos inicialmente dos $N = 2^{15}$ vértices do grafo e estabelecemos que a probabilidade de um vértice se conectar a outro seja p . A cada passo de tempo um vértice será conectado a outro com probabilidade p , até que sejam tomados todos os possíveis pares de vértices. Como estamos interessados em construir grafos aleatórios com determinadas conectividades médias por vértice (z), tomaremos a probabilidade p em função de z . Nesse caso, como esperamos que sejam construídas z dentre as possíveis $N - 1$ arestas por vértice, temos que $p = \frac{z}{N-1}$.

Um pequeno problema surge na construção dos grafos para pequenos valores de z . Para pequenas conectividades médias, aumenta a probabilidade de ocorrência de aglomerados de tamanho pequeno, ou seja, subgrafos isolados cujos nós não estão conectados aos demais nós do grafo (veja a figura 4.1). Para a realização das caminhadas, esses aglomerados devem ser eliminados do grafo principal. Em nossas simulações, para os valores de z iguais a 2, 4 e 6, após a eliminação dos aglomerados, permanecemos com 81,7%, 98,2% e 99,7% dos nós, respectivamente. Essa diminuição no número de nós para pequenas conectividades não é relevante em nosso estudo, visto que o percentual de vértices eliminados é pequeno. Para valores de z maiores do que 6, o percentual de aproveitamento foi de 100%.

4.2 A atribuição dos fitness

Para cada rede (livre de escala ou grafo aleatório) atribuímos os fitness a cada nó (vértice) utilizando as distribuições:

1. Normal
2. Exponencial bivariada.

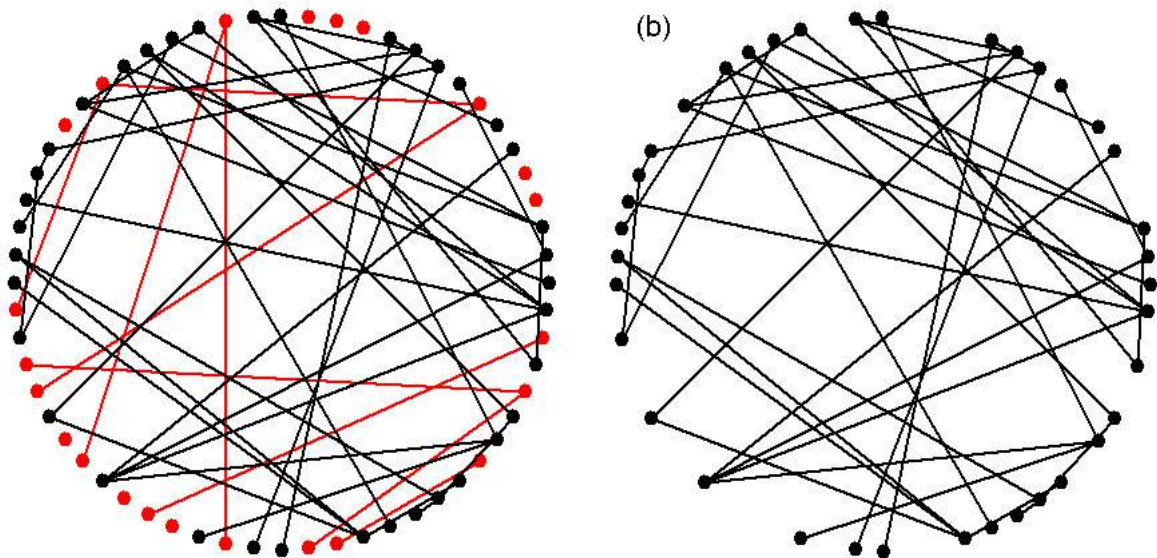


Figura 4.1: Em (a) um grafo aleatório com os aglomerados de tamanho pequeno em destaque. Em (b), o grafo obtido após sua eliminação.

Escolhemos essas duas distribuições para estudar as propriedades estatísticas das caminhadas aleatórias sobre relevos onde a correlação entre os valores adaptativos dos nós é positiva (distribuição normal) e negativa (distribuição exponencial bivariada).

4.2.1 Distribuição normal

A forma de atribuição dos fitness foi a mesma nas redes livres de escala e nos grafos aleatórios. Na atribuição dos fitness em cada rede através da distribuição normal o procedimento foi o seguinte:

- Resultados teóricos e experimentais envolvendo os efeitos de mutações benéficas sobre a evolução adaptativa de populações microbiais mostram que a distribuição de probabilidade dos efeitos de fitness de mutações benéficas é exponencial (Rozen et al, 2002). Baseados nos resultados obtidos nesses estudos, sorteamos o primeiro nó e a ele atribuímos um fitness sorteando um número (S_1) através da distribuição exponencial com média $1/20$;
- Para cada vizinho i desse nó, sorteia-se um fitness S_i através da distribuição normal com média λS_1 e variância $10^{-3}(1 - \lambda^2)$, em que λ representa o parâmetro de correlação entre os fitness;
- Para cada um desses nós aos quais foram atribuídos o fitness S_i , verificamos seus vizinhos mais próximos (j) aos quais ainda não foram atribuídos fitness e novamente

sorteamos um fitness S_j através da distribuição normal com média λS_i e variância $10^{-3}(1 - \lambda^2)$;

- Repete-se o procedimento até que tenham sido atribuídos fitness a todos os nós;
- Finalmente, para cada nó i com fitness S_i , obtemos o fitness definitivo como sendo $F_i = 1 + S_i$, para garantir que todos os valores sejam positivos.

4.2.2 Distribuição exponencial bivariada

A segunda distribuição que utilizamos na atribuição dos fitness foi a exponencial bivariada (Gumbel, 1960). Nesse modelo, o fitness de cada nó é atribuído a partir do fitness de um vizinho pela distribuição condicional de $Y = S_j$ dado $X = S_i$ cuja função densidade de probabilidade é dada por

$$f(S_j) = \alpha e^{-\alpha S_j(1+\alpha\theta S_i)} [(1 + \alpha\theta S_i)(1 + \alpha\theta S_j) - \theta],$$

onde $S_j > 0$, $S_i > 0$ e $0 \leq \theta \leq 1$ é o parâmetro de correlação. O procedimento computacional foi o seguinte:

- Um primeiro nó é sorteado e a ele atribui-se um fitness sorteando um número (S_1) através da distribuição exponencial com parâmetro $\alpha = 20$, ou seja, com média $1/\alpha = 1/20$;
- Para cada vizinho i desse nó, sorteia-se um fitness S_i através da distribuição exponencial bivariada com os parâmetros α e θ conforme descrito anteriormente. Nesse caso estamos tomando os valores S_i correlacionados com o valor S_1 ;
- Para cada um desses nós aos quais foram atribuídos o fitness S_i , verificamos seus vizinhos mais próximos (j) aos quais ainda não foram atribuídos fitness e novamente sorteamos um fitness S_j através da distribuição exponencial bivariada, da mesma maneira como no item anterior;
- Repete-se o procedimento até que tenham sido atribuídos fitness a todos os nós.

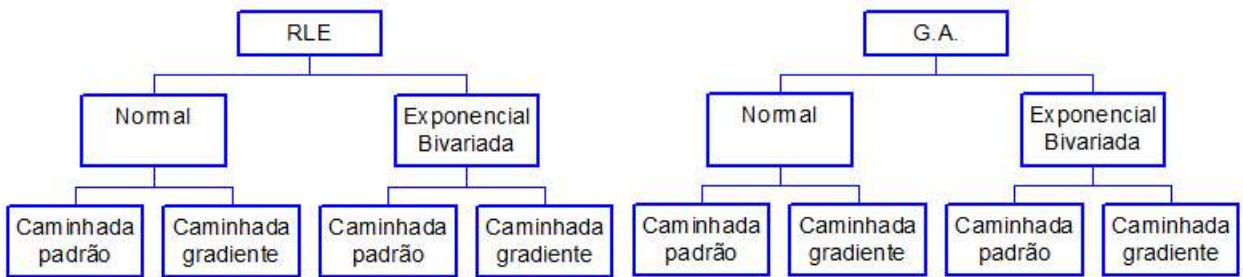


Figura 4.2: Organograma das simulações

4.3 As propriedades estudadas

Para uma melhor visualização das simulações realizadas, podemos observar o organograma apresentado na Figura 4.2. Foram construídas 1000 redes livres de escala e 1000 grafos aleatórios, atribuindo-lhes as duas distribuições de fitness citadas anteriormente. Foram realizadas caminhadas padrão e gradiente para cada rede e, para cada valor de z , foram calculadas as seguintes propriedades:

- **Comprimento médio das caminhadas não-nulas**

Para cada valor de z , calculamos o comprimento médio \bar{L} das caminhadas padrão e gradiente com comprimentos não nulos.

- **Densidade média de ótimos locais**

Para cada rede, seja N o número total de nós e n o número de ótimos locais. Calculamos a fração η de nós que são ótimos locais, ou seja,

$$\eta = \frac{n}{N},$$

e finalmente, para cada valor de z calculamos a média $\bar{\eta}$ entre os valores de η .

- **Diferença relativa entre os fitness local e global**

Para cada rede, calculamos o valor de

$$\Delta F_G = \frac{F_{Global} - F_{Local}}{F_{Global}},$$

onde F_{Global} é o valor máximo de fitness na rede e F_{Local} é o valor médio dos fitness dos ótimos locais das caminhadas. Finalmente, para cada valor de z , calculamos a média $\overline{\Delta F_G}$ entre os valores de ΔF_G .

- **Distribuição de frequências dos comprimentos das caminhadas**

Para cada valor de z , calculamos a fração média de caminhadas não-nulas com comprimento L ($L > 0$).

- **Rank das caminhadas**

Para cada rede, os fitness dos 1000 nós foram colocados em ordem decrescente, e nessa ordem os nós foram numerados de 1 a 1000. Para cada valor de z , calculamos o comprimento médio das caminhadas que partem dos nós de mesma numeração.

4.4 Resultados das caminhadas

4.4.1 Comprimento médio das caminhadas não-nulas

Observando os resultados dos comprimentos médios das caminhadas padrão como função da conectividade (Figura 4.3), percebemos que não houve diferenças significativas em relação às diferentes topologias consideradas, bem como em relação ao fitness com distribuição normal e com distribuição exponencial bivariada. Também não houve grandes variações nos resultados quando variamos o parâmetro de correlação λ . Para todos os casos de caminhada padrão, percebemos que o comprimento médio \bar{L} cresce com z , com uma taxa de crescimento maior para pequenos valores de z e diminuindo essa taxa de crescimento para valores maiores de z . As barras de erro não aparecem pelo fato de serem muito pequenas.

De acordo com os resultados apresentados na Figura 4.4, em relação às caminhadas gradiente, o comprimento médio das caminhadas sofreu um crescimento acentuado para pequenas conectividades, estabilizando-se a partir de um valor próximo de $z = 14$. Observamos que o número médio de passos é menor do que o observado nas caminhadas padrão. Quando $\lambda > 0$ (relevo correlacionado), os comprimentos médios das caminhadas apresentaram um comportamento qualitativo diferente do obtido nas caminhadas padrão. Verificamos nesse caso, um crescimento mais acentuado de \bar{L} para pequenos valores de z , atingindo um pico para um valor de z também pequeno ($z = 4$), a partir daí decrescendo e assumindo um valor próximo do obtido para relevo não correlacionado, para valores maiores de z . Interessante verificar que mais uma vez o comportamento qualitativo não mudou para diferentes topologias e distribuições do fitness. Constatamos que, para $\lambda = 0$, o valor de \bar{L} se aproxima do valor estimado por Orr (2003) em $\bar{L} = e - 1 \approx 1,72$, discutido no capítulo anterior. Percebemos também que, para valores pequenos de z , os comprimentos médios foram um pouco maiores para a distribuição normal. As barras de erro não aparecem pelo fato de serem muito pequenas.

4.4.2 Densidade média de ótimos locais

Estudando a densidade de ótimos locais nas redes ($\bar{\eta}$), observamos um mesmo comportamento para as duas topologias consideradas (as redes livres de escala e os grafos

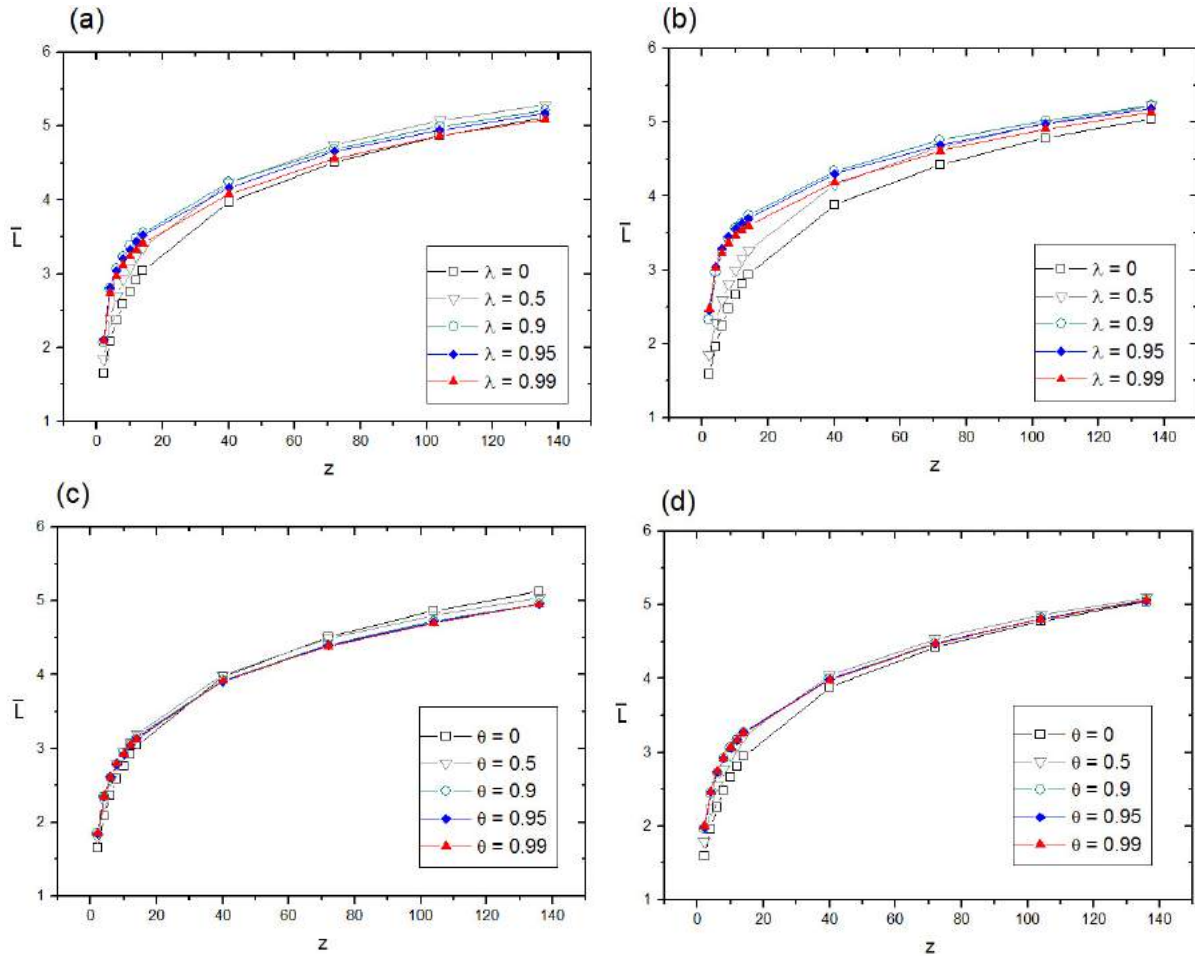


Figura 4.3: Comprimento médio das caminhadas padrão. Em (a) e (b) temos os resultados com a distribuição de fitness normal, para as redes livres de escala e os grafos aleatórios, respectivamente. Em (c) e (d) temos os resultados com a distribuição de fitness exponencial bivariada, para as redes livres de escala e os grafos aleatórios, respectivamente.

aleatórios), bem como para as diferentes distribuições de fitness (veja Figura 4.5). As barras de erro não aparecem pelo fato de serem muito pequenas. Em todos os casos observamos que o valor de $\bar{\eta}$ diminuiu com z de acordo com uma lei de potência. Em outras palavras, $\bar{\eta} \propto z^{-\alpha}$, com α próximo de 0.9 nas redes livres de escala e nos grafos aleatórios. Interessante observar que os resultados estão muito próximos dos obtidos por Campos e Moreira (2005), estudados na capítulo anterior. A tabela 1 contém os valores de α para cada tipo de rede e relevo de fitness, para as diferentes correlações.

4.4.3 Diferença relativa entre os fitness local e global ΔF_G

Como vimos na Seção 4.3, calculamos a diferença relativa entre os valores adaptativos dos mínimos locais e global, ΔF_G , fazendo $\Delta F_G = \frac{F_{Global} - F_{Local}}{F_{Global}}$. Um valor pequeno

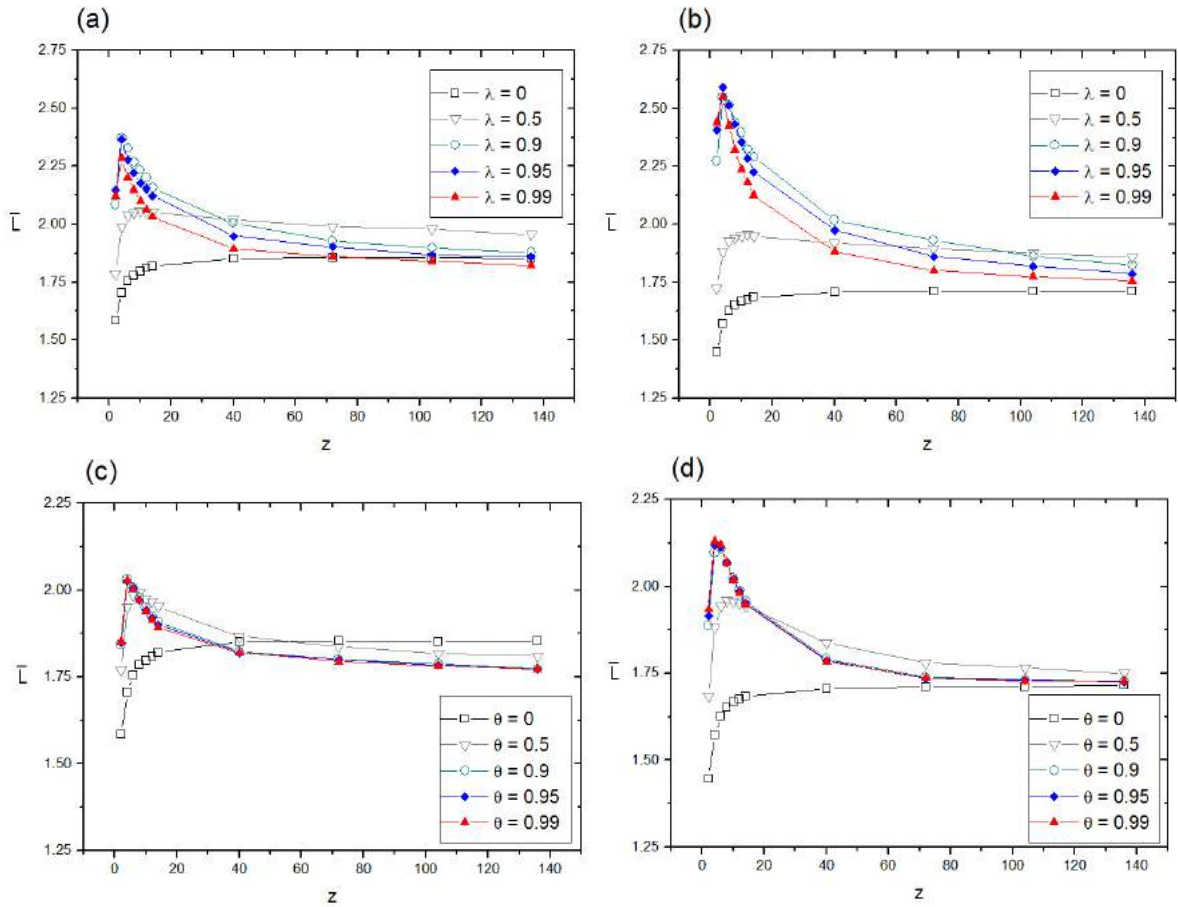


Figura 4.4: Comprimento médio das caminhadas gradiente. Em (a) e (b) temos os resultados com a distribuição de fitness normal, para as redes livres de escala e os grafos aleatórios, respectivamente. Em (c) e (d) temos os resultados com a distribuição de fitness exponencial bivariada, para as redes livres de escala e os grafos aleatórios, respectivamente.

de ΔF_G indica que as intensidades dos mínimos locais encontrados nas caminhadas estão próximas à intensidade do ótimo global.

Nas Figuras 4.6 e 4.7, observamos que $\overline{\Delta F_G}$ decresce com z de acordo com uma lei de potência, independentemente da topologia considerada ou da distribuição do fitness. Em outras palavras, $\overline{\Delta F_G} \propto z^{-\beta}$, com β próximo de 0.2 nas redes livres de escala e nos grafos aleatórios. As tabelas 2 e 3 contêm os valores obtidos para β para relevo de fitness normal e exponencial, respectivamente. Verificamos que $\overline{\Delta F_G}$ diminui com a correlação quando a distribuição normal é considerada. Um comportamento oposto é observado para a distribuição exponencial.

Tabela 1: Valores de α obtidos nas caminhadas.

Correlação	α (RLE/Normal)	α (GA/Normal)	α (RLE/Exponencial)	α (GA/Exponencial)
0	$0,92 \pm 0,02$	$0,95 \pm 0,02$	$0,92 \pm 0,02$	$0,95 \pm 0,02$
0,5	$0,93 \pm 0,01$	$0,94 \pm 0,02$	$0,88 \pm 0,01$	$0,93 \pm 0,02$
0,9	$0,85 \pm 0,01$	$0,87 \pm 0,01$	$0,86 \pm 0,01$	$0,91 \pm 0,01$
0,95	$0,84 \pm 0,01$	$0,86 \pm 0,01$	$0,86 \pm 0,01$	$0,90 \pm 0,01$
0,99	$0,84 \pm 0,01$	$0,86 \pm 0,01$	$0,86 \pm 0,01$	$0,90 \pm 0,01$

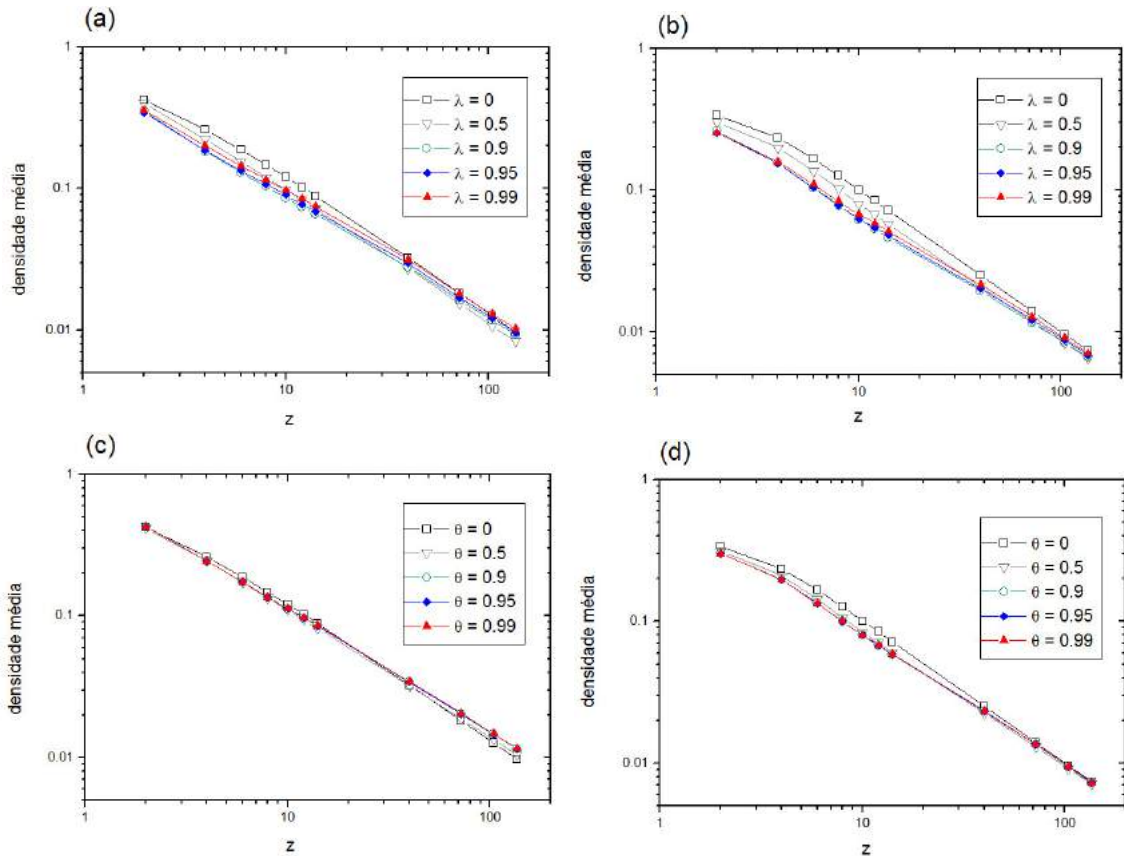


Figura 4.5: Densidade média de ótimos locais nas redes. Em (a) e (b) temos os resultados com a distribuição de fitness normal, para as redes livres de escala e os grafos aleatórios, respectivamente. Em (c) e (d) temos os resultados com a distribuição de fitness exponencial bivariada, para as redes livres de escala e os grafos aleatórios, respectivamente.

Fizemos um comparativo gráfico entre as caminhadas padrão e gradiente a fim de observar qual o tipo de caminhada mais eficiente, ou seja, qual das caminhadas apresenta menores valores de $\overline{\Delta F_G}$, indicando que as mesmas levaram a ótimos locais mais próximos do ótimo global. Podemos constatar, através das Figuras 4.8 e 4.9, que a caminhada gradiente é mais eficiente do que a caminhada padrão, o que já era esperado, já que na caminhada gradiente cada passo é dado para o vizinho com maior valor adaptativo.

Tabela 2: Valores de β obtidos nas caminhadas com distribuição de fitness normal.

Correlação	β (RLE/Padrão)	β (RLE/Gradiente)	β (GA/Padrão)	β (GA/Gradiente)
0	$0,14 \pm 0,01$	$0,16 \pm 0,01$	$0,14 \pm 0,01$	$0,15 \pm 0,01$
0,5	$0,21 \pm 0,01$	$0,23 \pm 0,01$	$0,20 \pm 0,01$	$0,22 \pm 0,01$
0,9	$0,26 \pm 0,01$	$0,27 \pm 0,02$	$0,28 \pm 0,01$	$0,29 \pm 0,01$
0,95	$0,27 \pm 0,01$	$0,29 \pm 0,02$	$0,30 \pm 0,01$	$0,32 \pm 0,01$
0,99	$0,30 \pm 0,01$	$0,31 \pm 0,01$	$0,34 \pm 0,02$	$0,35 \pm 0,02$

Tabela 3: Valores de β obtidos nas caminhadas com distribuição de fitness exponencial bivariada.

Correlação	β (RLE/Padrão)	β (RLE/Gradiente)	β (GA/Padrão)	β (GA/Gradiente)
0	$0,10 \pm 0,01$	$0,12 \pm 0,01$	$0,09 \pm 0,01$	$0,10 \pm 0,01$
0,5	$0,15 \pm 0,01$	$0,17 \pm 0,01$	$0,13 \pm 0,01$	$0,16 \pm 0,01$
0,9	$0,15 \pm 0,01$	$0,17 \pm 0,01$	$0,13 \pm 0,01$	$0,15 \pm 0,01$
0,95	$0,15 \pm 0,01$	$0,17 \pm 0,01$	$0,13 \pm 0,01$	$0,14 \pm 0,01$
0,99	$0,15 \pm 0,01$	$0,17 \pm 0,01$	$0,13 \pm 0,01$	$0,14 \pm 0,01$

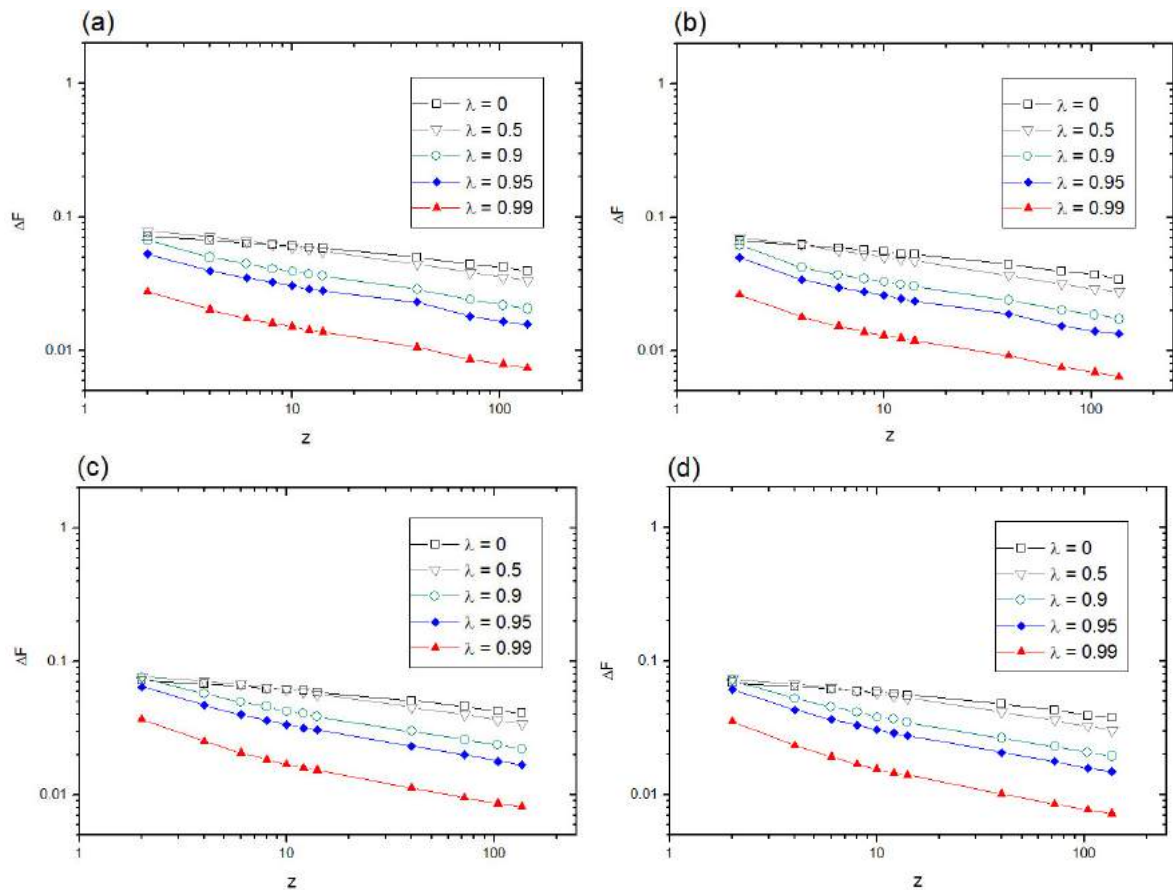


Figura 4.6: Resultados de $\overline{\Delta F}_G$ em função de z , com a distribuição de fitness normal. Em (a) e (b) temos os resultados nas redes livres de escala, para as caminhadas padrão e gradiente, respectivamente. Em (c) e (d) temos os resultados nos grafos aleatórios, para as caminhadas padrão e gradiente, respectivamente.

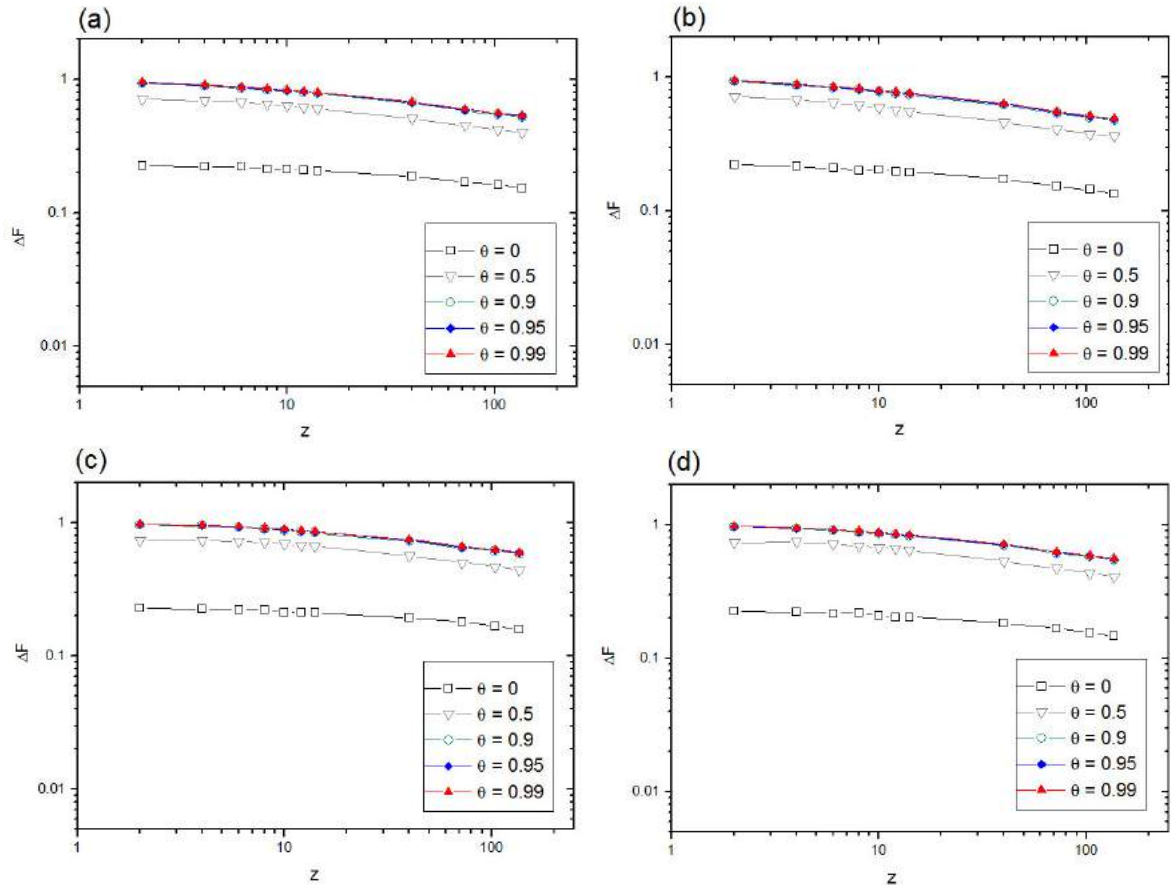


Figura 4.7: Resultados de $\overline{\Delta F_G}$ em função de z , com a distribuição de fitness exponencial bivariada. Em (a) e (b) temos os resultados nas redes livres de escala, para as caminhadas padrão e gradiente, respectivamente. Em (c) e (d) temos os resultados nos grafos aleatórios, para as caminhadas padrão e gradiente, respectivamente.

4.4.4 Distribuição de frequências dos comprimentos das caminhadas

Com relação à distribuição de frequência dos comprimentos das caminhadas, não observamos diferenças importantes para as diferentes topologias ou para as diferentes distribuições de fitness. Foram observadas diferenças qualitativas apenas para os dois tipos diferentes de caminhadas realizadas (padrão e gradiente). Também não foram percebidas diferenças significativas para os diferentes parâmetros de correlação.

Nas caminhadas padrão (veja figuras 4.10, 4.11, 4.12 e 4.13), observamos que para valores pequenos de z temos uma maior frequência de caminhadas com comprimentos 1 e 2, com uma frequência cada vez menor para comprimentos maiores do que 2. Para conectividades grandes observamos uma frequência pequena de caminhadas com comprimento pequeno, o que é esperado, já que a densidade de ótimos locais diminui para conectividades grandes. Essa frequência cresce com o comprimento da caminhada, atinge um máximo e volta a decrescer.

Nas caminhadas gradiente (veja figuras 4.14, 4.15, 4.16 e 4.17), observamos que a frequência de caminhadas de comprimento pequeno é muito grande para valores pequenos de z , decrescendo e assumindo valores muito pequenos para valores grandes de z . Caminhadas com comprimentos 1 e 2 possuem grandes frequências para todos os valores de z . A frequência decresce rapidamente para comprimentos maiores do que 2, sendo praticamente nula para comprimentos maiores do que 6.

4.4.5 Rank das caminhadas

Procuramos, agora, entender se os valores dos fitness dos nós em que a caminhada se inicia é importante na dinâmica da caminhada. Por esta razão, estudamos a dependência do comprimento da caminhada com o ranking do nó onde a caminhada é iniciada.

Para a caminhada padrão (veja figuras 4.18, 4.19, 4.20 e 4.21), não percebemos diferenças qualitativas importantes para as diferentes topologias e distribuições de fitness consideradas. Também não observamos diferenças para os diferentes valores de λ ($\lambda = 0, 0.5, 0.9, 0.95$ ou 0.99). Para todos esse valores de λ observamos que quanto maior a conectividade z , mais rapidamente cresce o valor de \bar{L} , de modo que para cada rank, maior conectividade sempre corresponde a um maior valor de \bar{L} . Percebemos que, para valores pequenos de z , o comprimento médio \bar{L} cresce de forma aproximadamente linear com o rank, e a partir de valores intermediários de z , \bar{L} tem um crescimento inicial abrupto e cresce de forma aproximadamente linear para valores intermediários e grandes do rank. Observamos também que, para esse tipo de caminhada, fixado o rank, sempre temos valores maiores de \bar{L} para valores maiores de z , ou seja, quanto maior a conectividade, maior será o comprimento da caminhada, para o mesmo rank. Tal comportamento não foi observado na caminhada gradiente, conforme veremos a seguir.

Para a caminhada gradiente (veja as figuras 4.22, 4.23, 4.24 e 4.25), observamos comportamentos diferentes para diferentes valores de λ , apesar de não percebermos diferenças quando modificamos a topologia ou a distribuição do fitness. Para $\lambda = 0$, e valores pequenos de z , o comportamento de \bar{L} se deu de forma aproximadamente linear com o rank. À medida em que aumentamos os valores de z , percebemos um crescimento inicial de \bar{L} cada vez mais brusco, estabilizando-se num valor máximo. O valor do rank para o

qual \bar{L} se torna independente do rank decresce com o aumento de z .

Com o aumento de λ observamos um crescimento mais rápido de \bar{L} com o rank. Verificamos que os comprimentos médios das caminhadas se tornam maiores para redes com valores pequenos de conectividade média, quando comparado com redes de grande conectividade, no regime de rank grande.

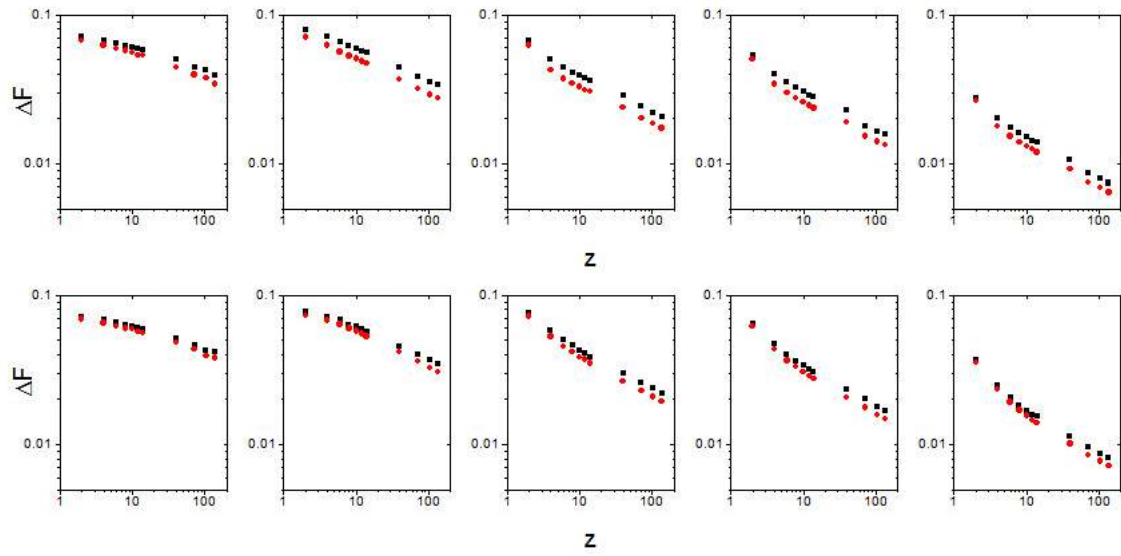


Figura 4.8: Gráficos comparativos dos valores de $\overline{\Delta F}_G$ em função de z , para as caminhadas padrão (quadrado preto) e gradiente (círculo vermelho), com distribuição normal de fitness. Nas linhas, de cima para baixo, temos respectivamente, as caminhadas nas redes livres de escala e nos grafos aleatórios. Nas colunas, da esquerda para a direita, temos, respectivamente, $\lambda = 0, 0.5, 0.9, 0.95$ e 0.99 .

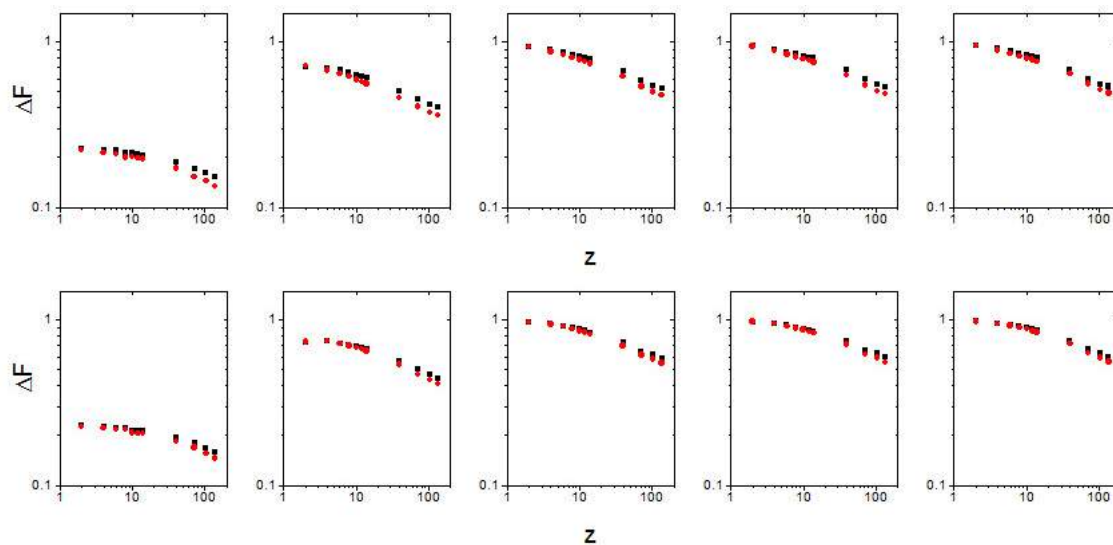


Figura 4.9: Gráficos comparativos dos valores de $\overline{\Delta F}_G$ em função de z , para as caminhadas padrão (quadrado preto) e gradiente (círculo vermelho), com distribuição exponencial de fitness. Nas linhas, de cima para baixo, temos respectivamente, as caminhadas nas redes livres de escala e nos grafos aleatórios. Nas colunas, da esquerda para a direita, temos, respectivamente, $\theta = 0, 0.5, 0.9, 0.95$ e 0.99 .

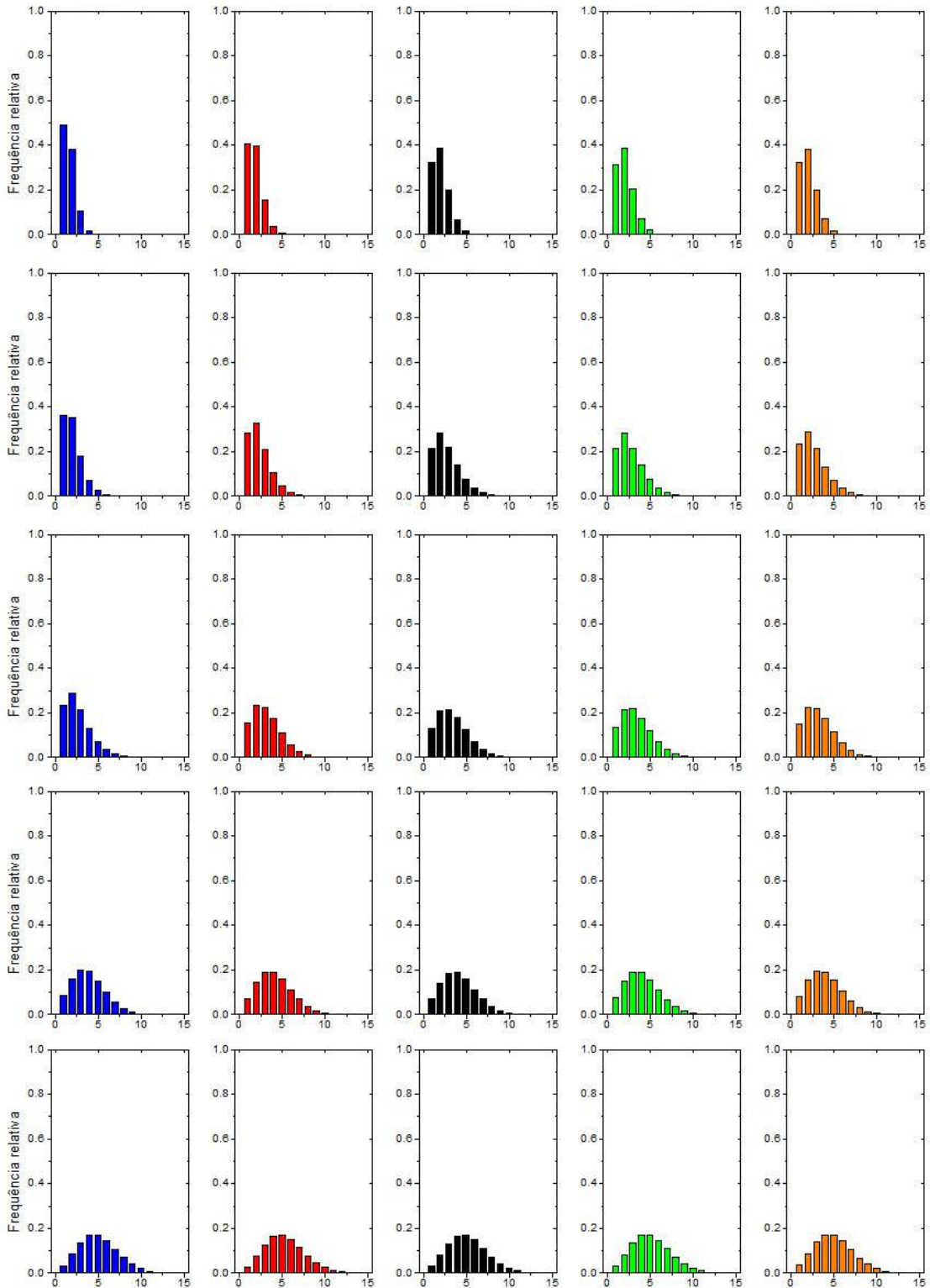


Figura 4.10: Distribuição de frequência dos números de passos das caminhadas padrão nas redes livres de escala, com distribuição normal de fitness. Considerando as linhas de cima para baixo, temos $z = 2, 4, 12, 40$ e 136 . Considerando as colunas da esquerda para a direita, temos $\lambda = 0, 0.5, 0.9, 0.95$ e 0.99 . Em cada gráfico, no eixo horizontal temos o número de passos das caminhadas.

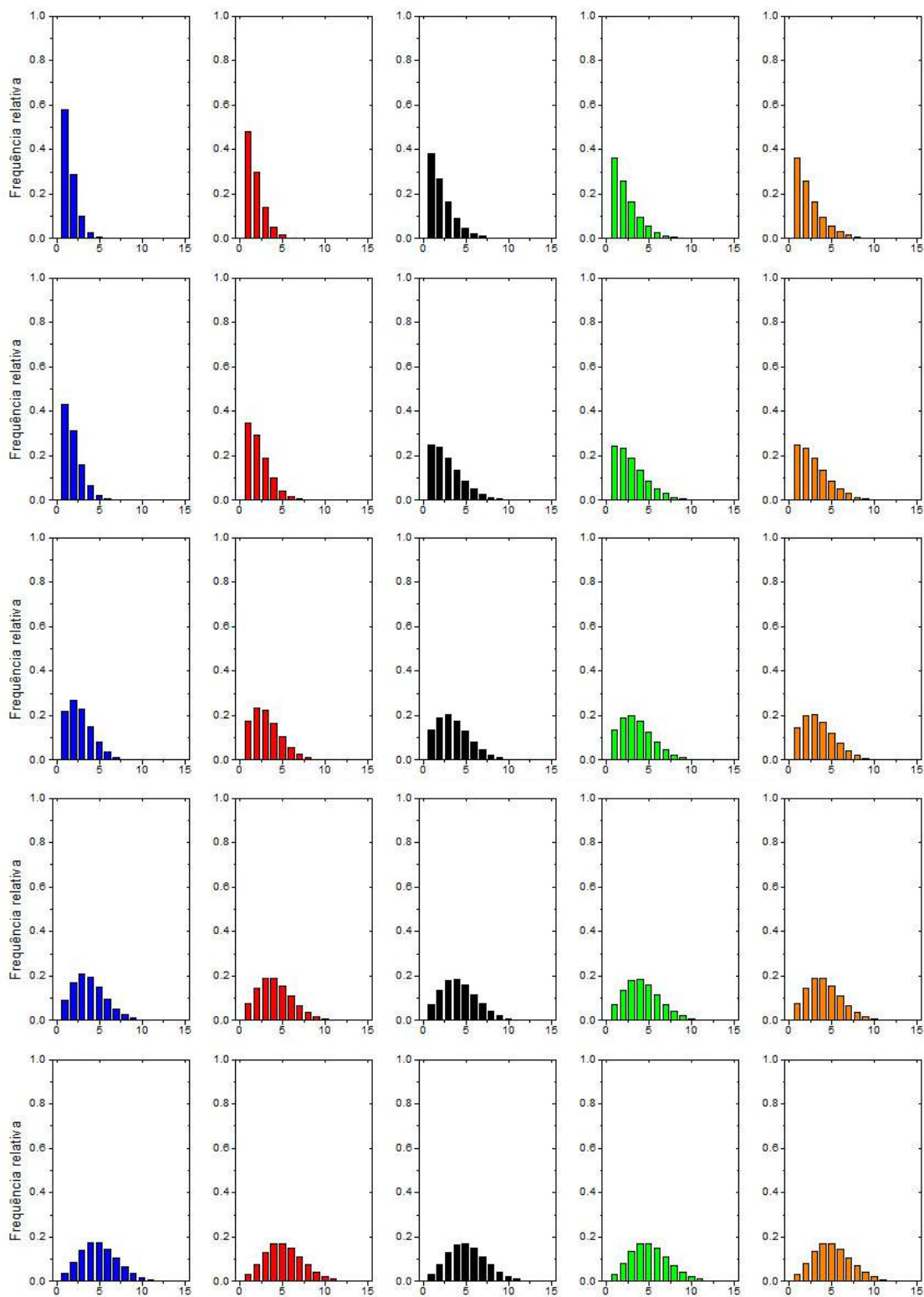


Figura 4.11: Distribuição de frequência dos números de passos das caminhadas padrão nos grafos aleatórios, com distribuição normal de fitness. Considerando as linhas de cima para baixo, temos $z = 2, 4, 12, 40$ e 136 . Considerando as colunas da esquerda para a direita, temos $\lambda = 0, 0.5, 0.9, 0.95$ e 0.99 . Em cada gráfico, no eixo horizontal temos o número de passos das caminhadas.

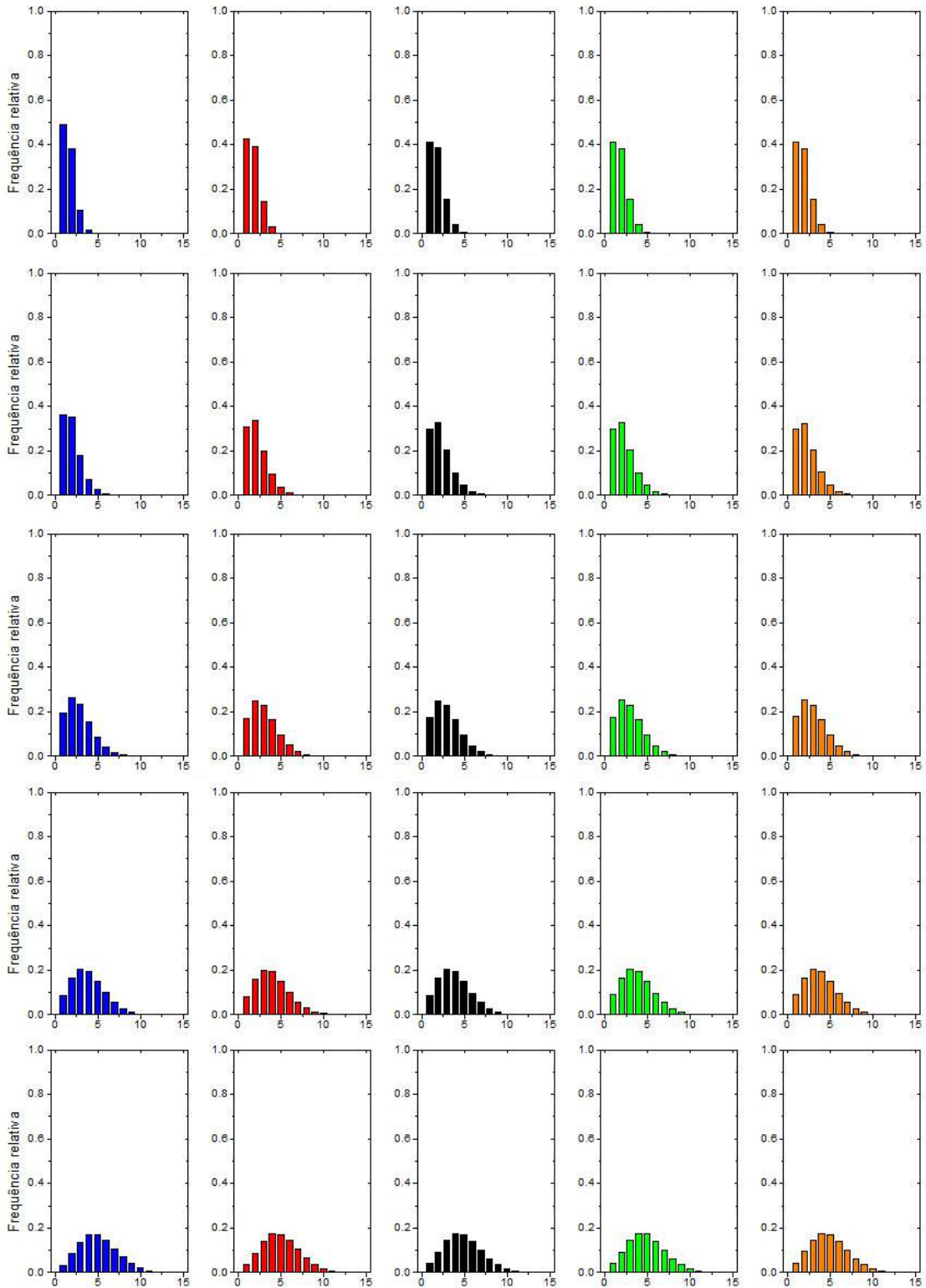


Figura 4.12: Distribuição de frequência dos números de passos das caminhadas padrão nas redes livres de escala, com distribuição de fitness exponencial biviariada. Considerando as linhas de cima para baixo, temos $z = 2, 4, 12, 40$ e 136 . Considerando as colunas da esquerda para a direita, temos $\theta = 0, 0.5, 0.9, 0.95$ e 0.99 . Em cada gráfico, no eixo horizontal temos o número de passos das caminhadas.

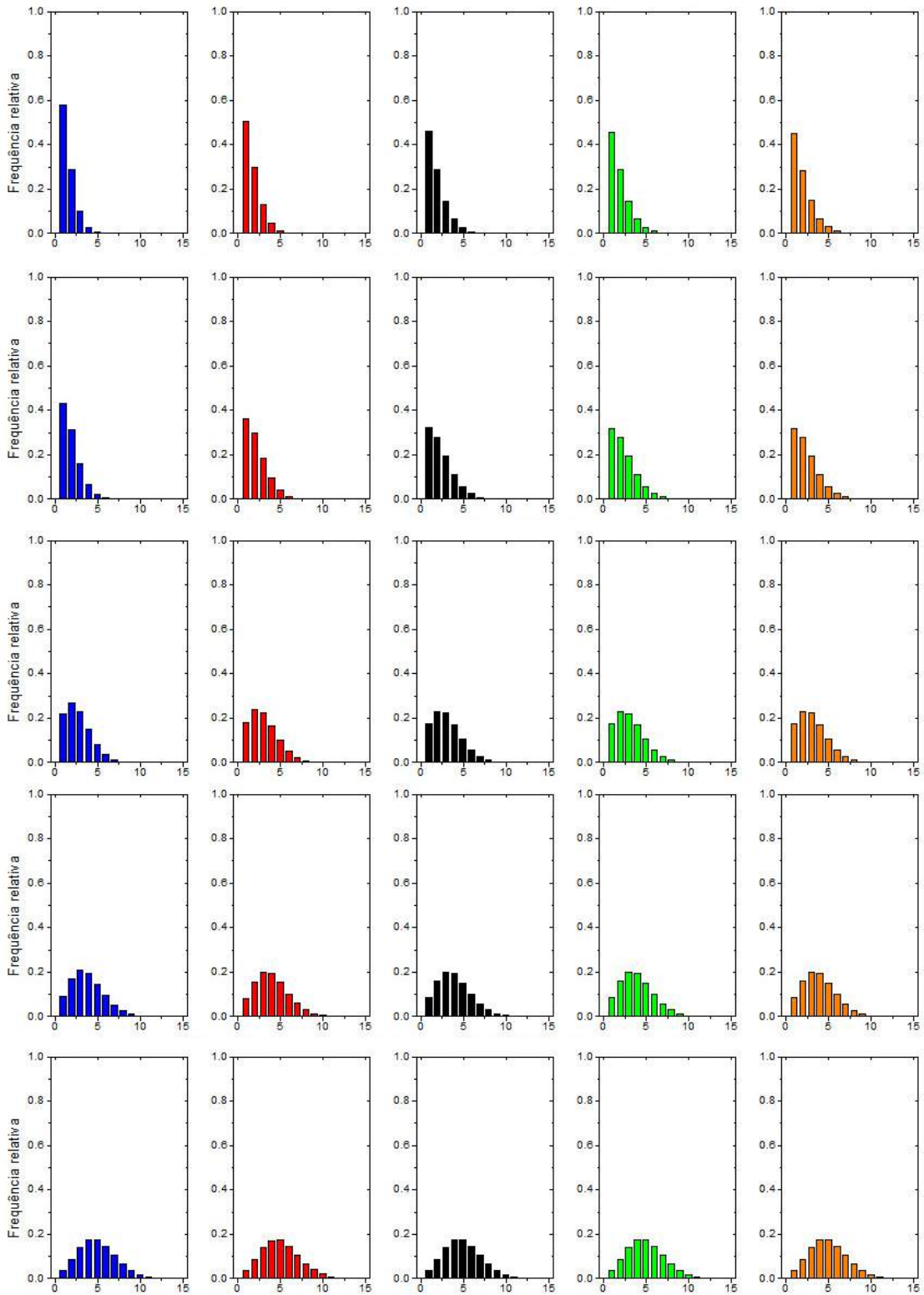


Figura 4.13: Distribuição de frequência dos números de passos das caminhadas padrão nos grafos aleatórios, com distribuição de fitness exponencial bivariada. Considerando as linhas de cima para baixo, temos $z = 2, 4, 12, 40$ e 136 . Considerando as colunas da esquerda para a direita, temos $\theta = 0, 0.5, 0.9, 0.95$ e 0.99 . Em cada gráfico, no eixo horizontal temos o número de passos das caminhadas.

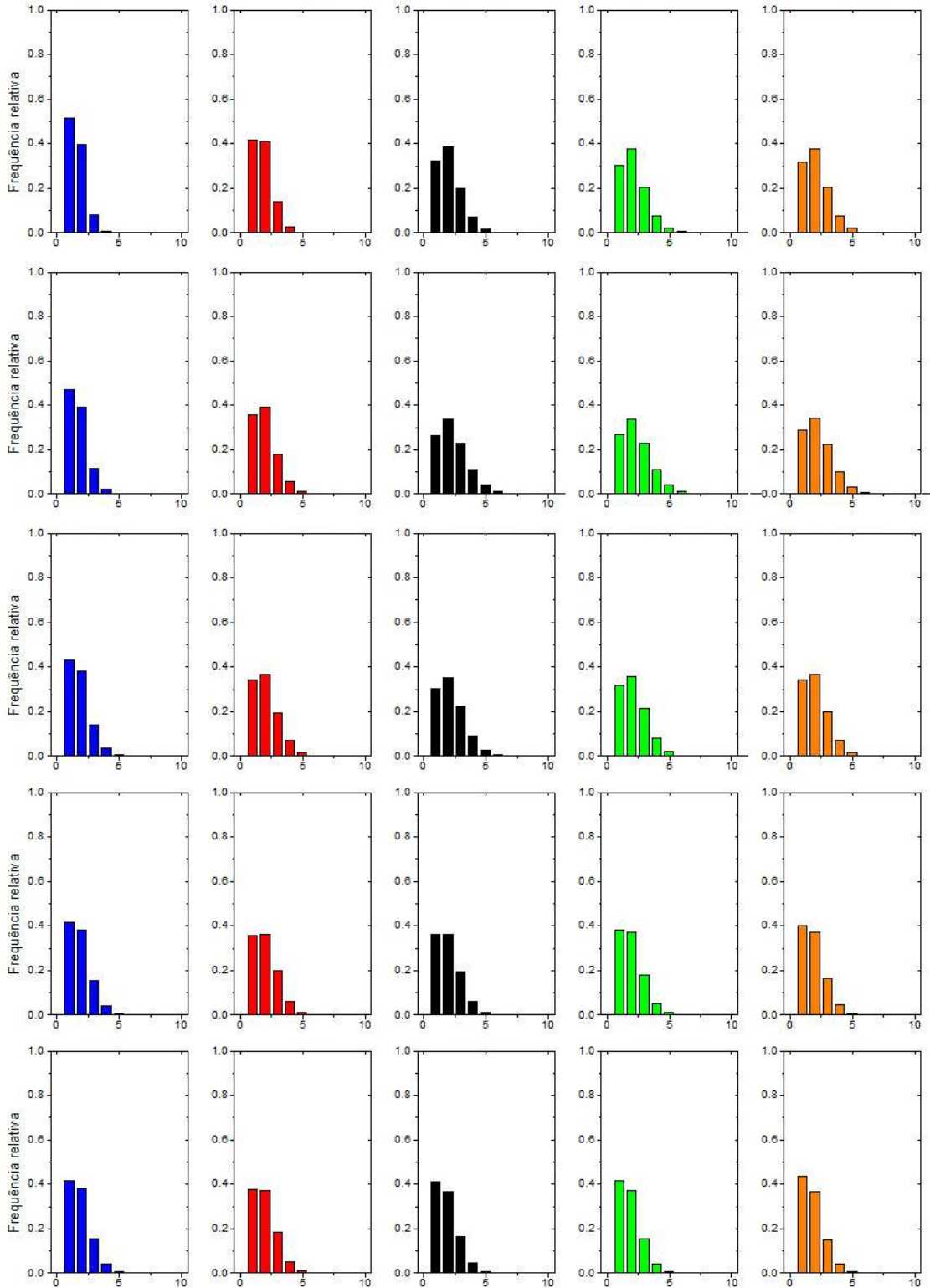


Figura 4.14: Distribuição de frequência dos números de passos das caminhadas gradiente nas redes livres de escala, com distribuição normal de fitness. Considerando as linhas de cima para baixo, temos $z = 2, 4, 12, 40$ e 136 . Considerando as colunas da esquerda para a direita, temos $\lambda = 0, 0.5, 0.9, 0.95$ e 0.99 . Em cada gráfico, no eixo horizontal temos o número de passos das caminhadas.

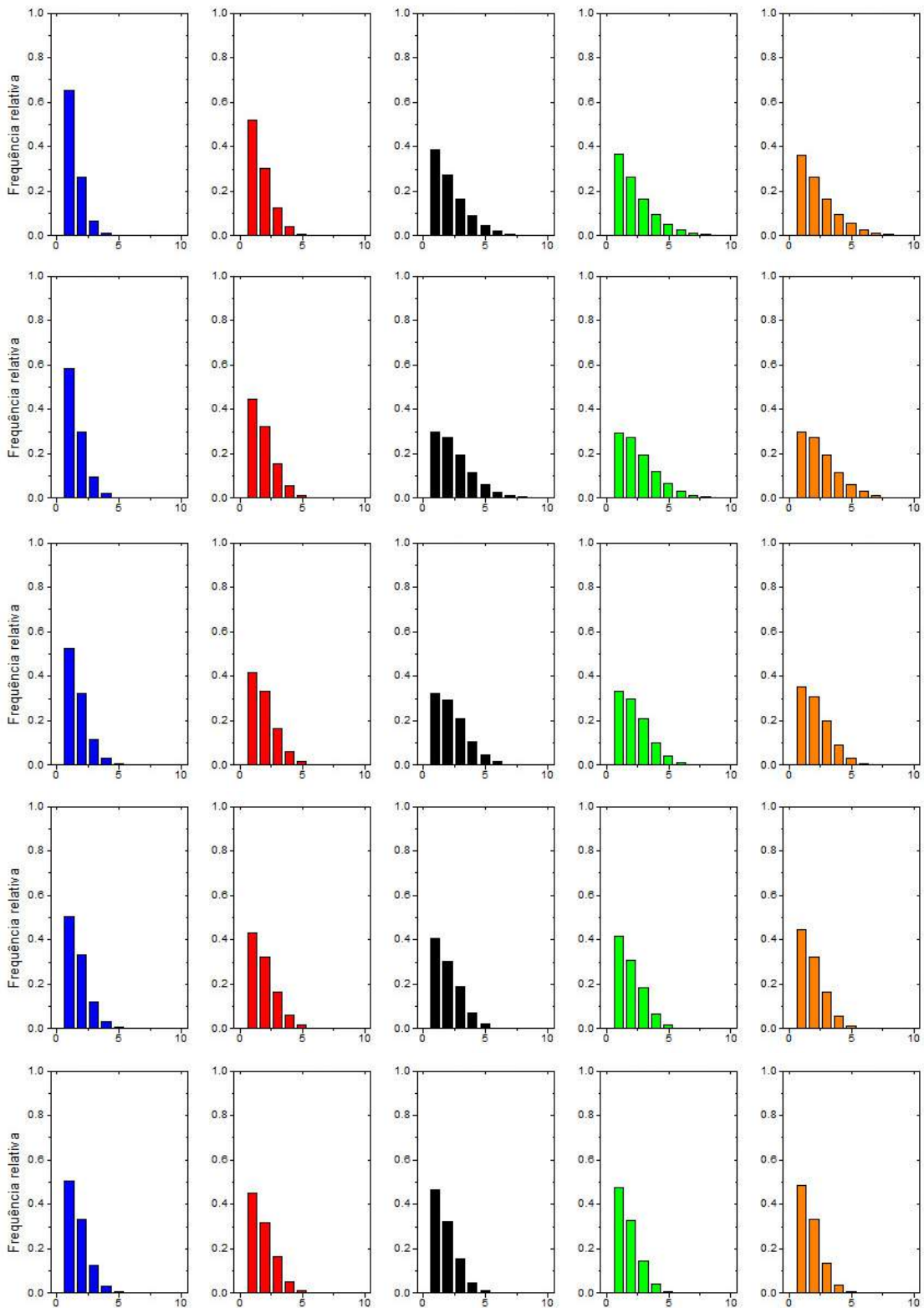


Figura 4.15: Distribuição de frequência dos números de passos das caminhadas gradiente nos grafos aleatórios, com distribuição normal de fitness. Considerando as linhas de cima para baixo, temos $z = 2, 4, 12, 40$ e 136 . Considerando as colunas da esquerda para a direita, temos $\lambda = 0, 0.5, 0.9, 0.95$ e 0.99 . Em cada gráfico, no eixo horizontal temos o número de passos das caminhadas.

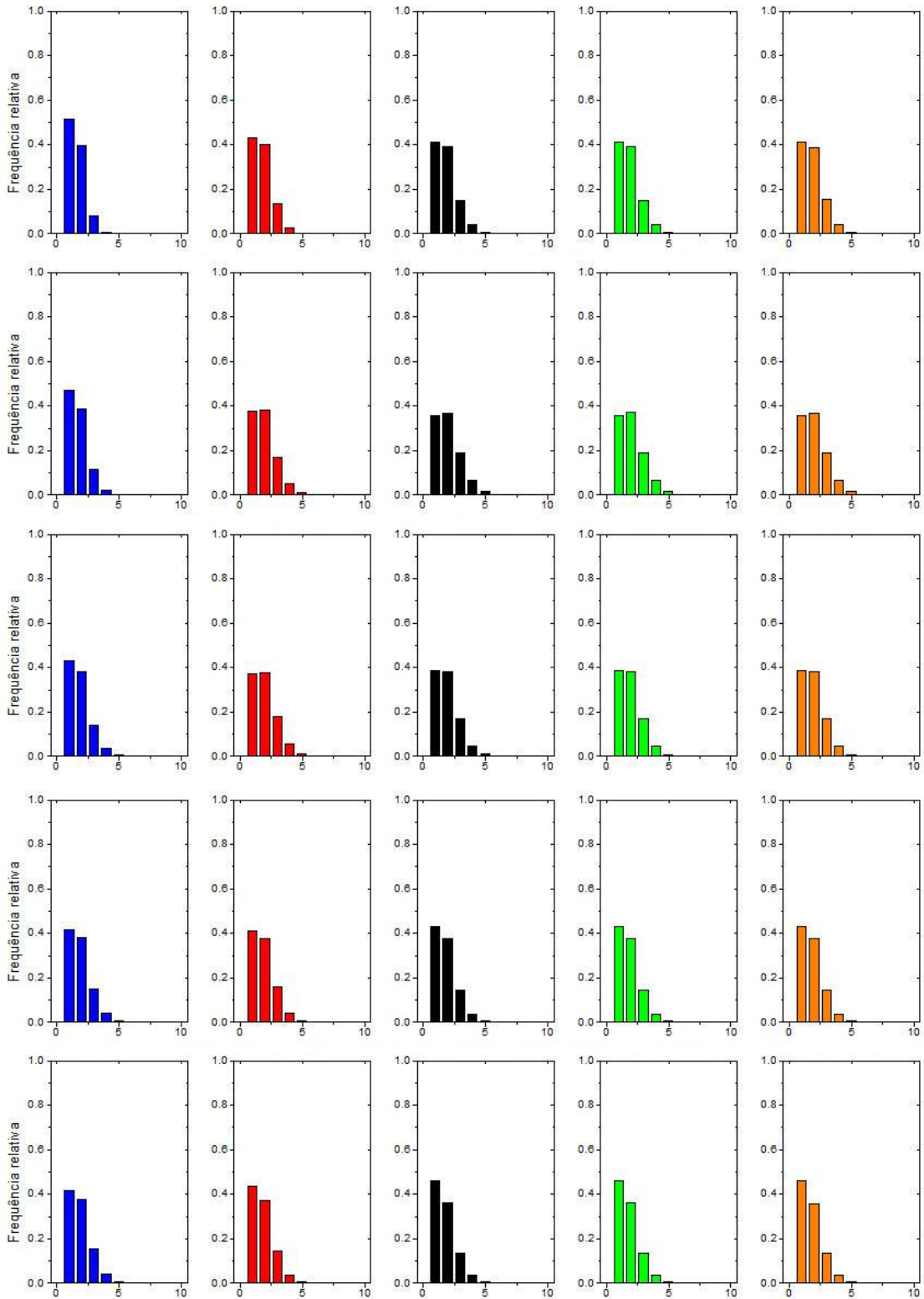


Figura 4.16: Distribuição de frequência dos números de passos das caminhadas gradiente nas redes livres de escala, com distribuição de fitness exponencial bivariada. Considerando as linhas de cima para baixo, temos $z = 2, 4, 12, 40$ e 136 . Considerando as colunas da esquerda para a direita, temos $\theta = 0, 0.5, 0.9, 0.95$ e 0.99 . Em cada gráfico, no eixo horizontal temos o número de passos das caminhadas.

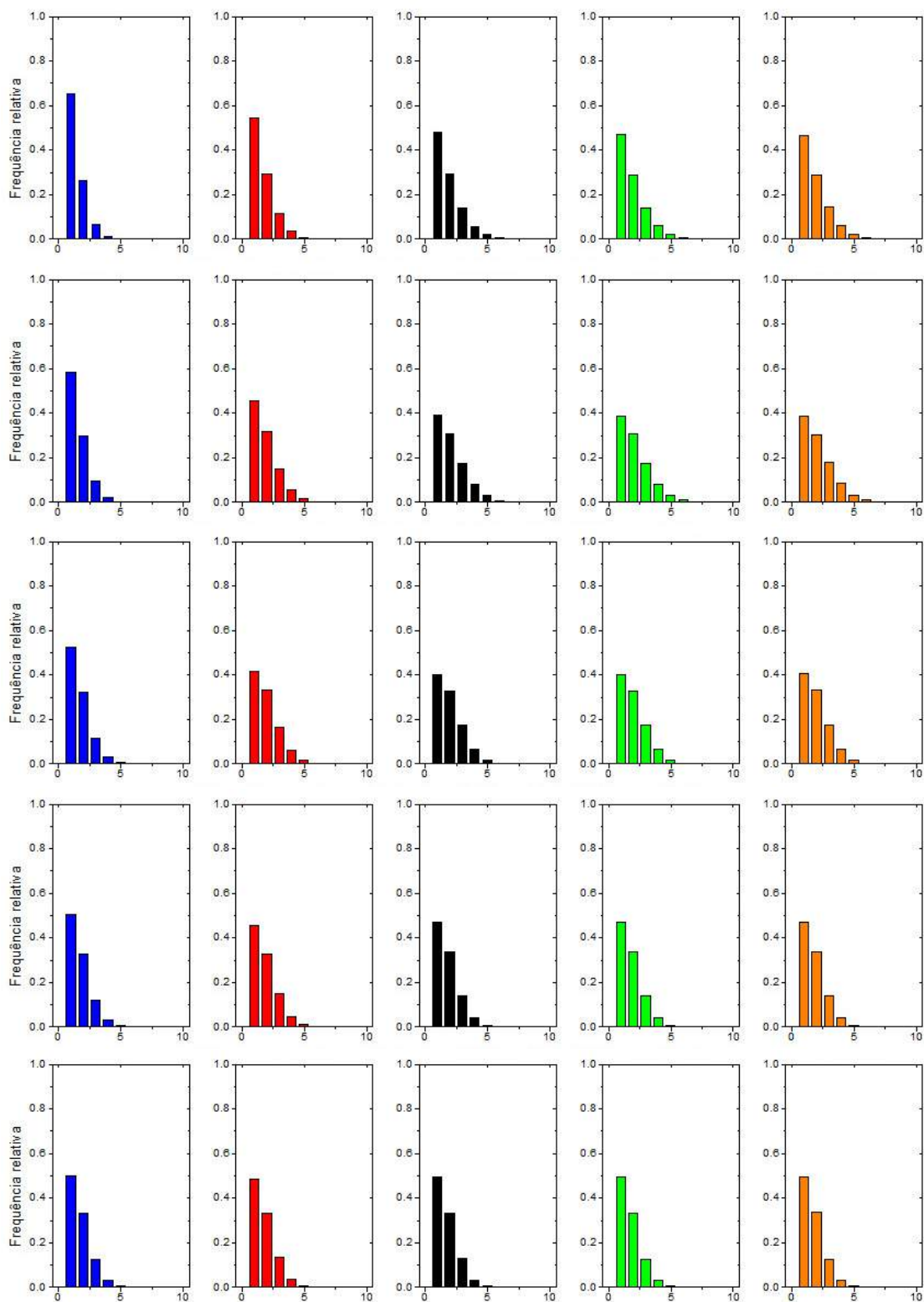


Figura 4.17: Distribuição de frequência dos números de passos das caminhadas gradiente nos grafos aleatórios, com distribuição de fitness exponencial bivariada. Considerando as linhas de cima para baixo, temos $z = 2, 4, 12, 40$ e 136 . Considerando as colunas da esquerda para a direita, temos $\theta = 0, 0.5, 0.9, 0.95$ e 0.99 . Em cada gráfico, no eixo horizontal temos o número de passos das caminhadas.

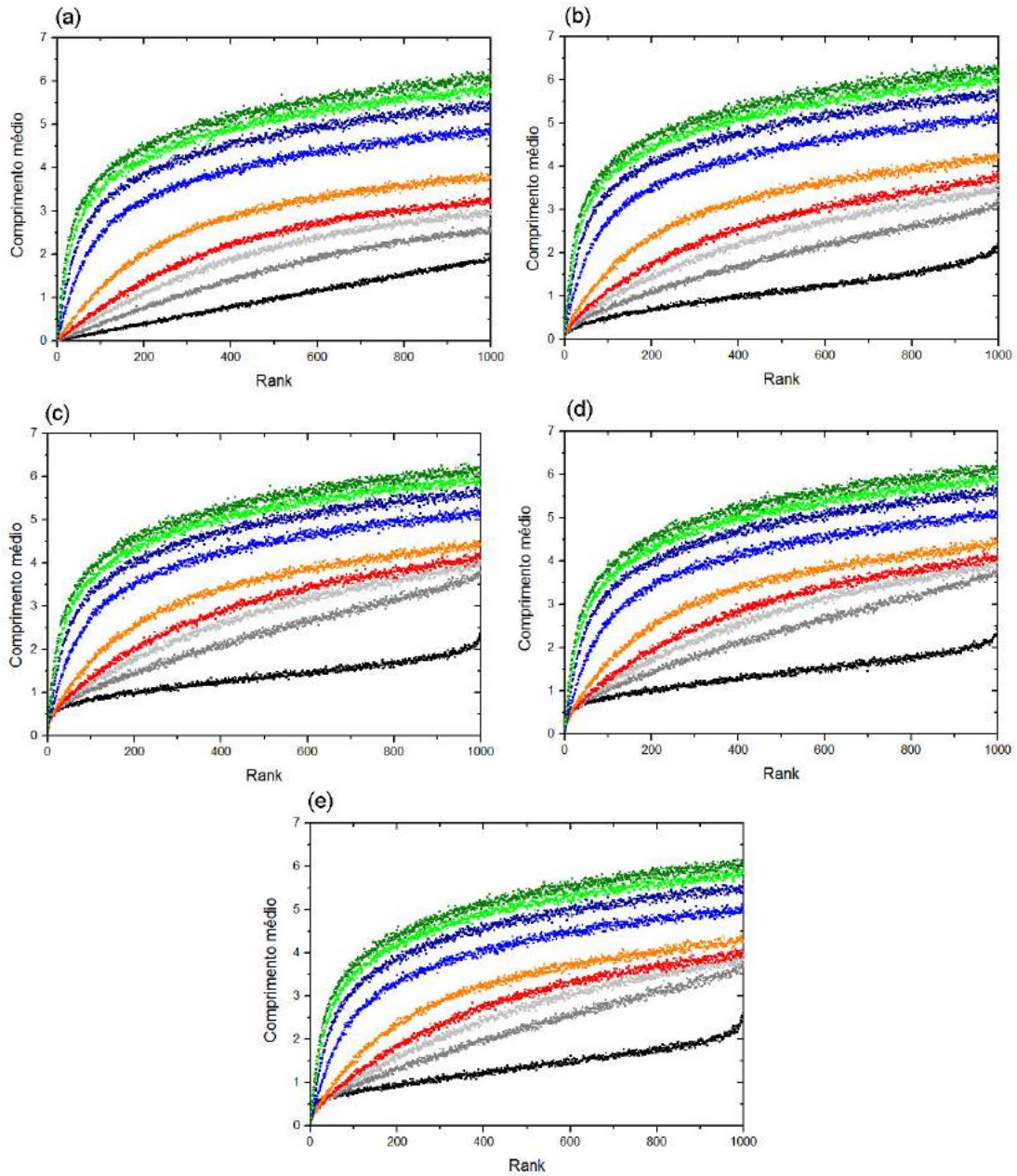


Figura 4.18: Rank das caminhadas padrão nas redes livres de escala, com distribuição normal de fitness. Para as figuras (a), (b), (c), (d) e (e) temos $\lambda = 0, 0.5, 0.9, 0.95$ e 0.99 , respectivamente. Em cada gráfico, para cada cor corresponde um valor de z , a saber: $z = 2$ (preto), $z = 4$ (cinza escuro), $z = 6$ (cinza claro), $z = 8$ (vermelho), $z = 14$ (laranja), $z = 40$ (azul claro), $z = 72$ (azul escuro), $z = 104$ (verde claro), $z = 136$ (verde escuro).

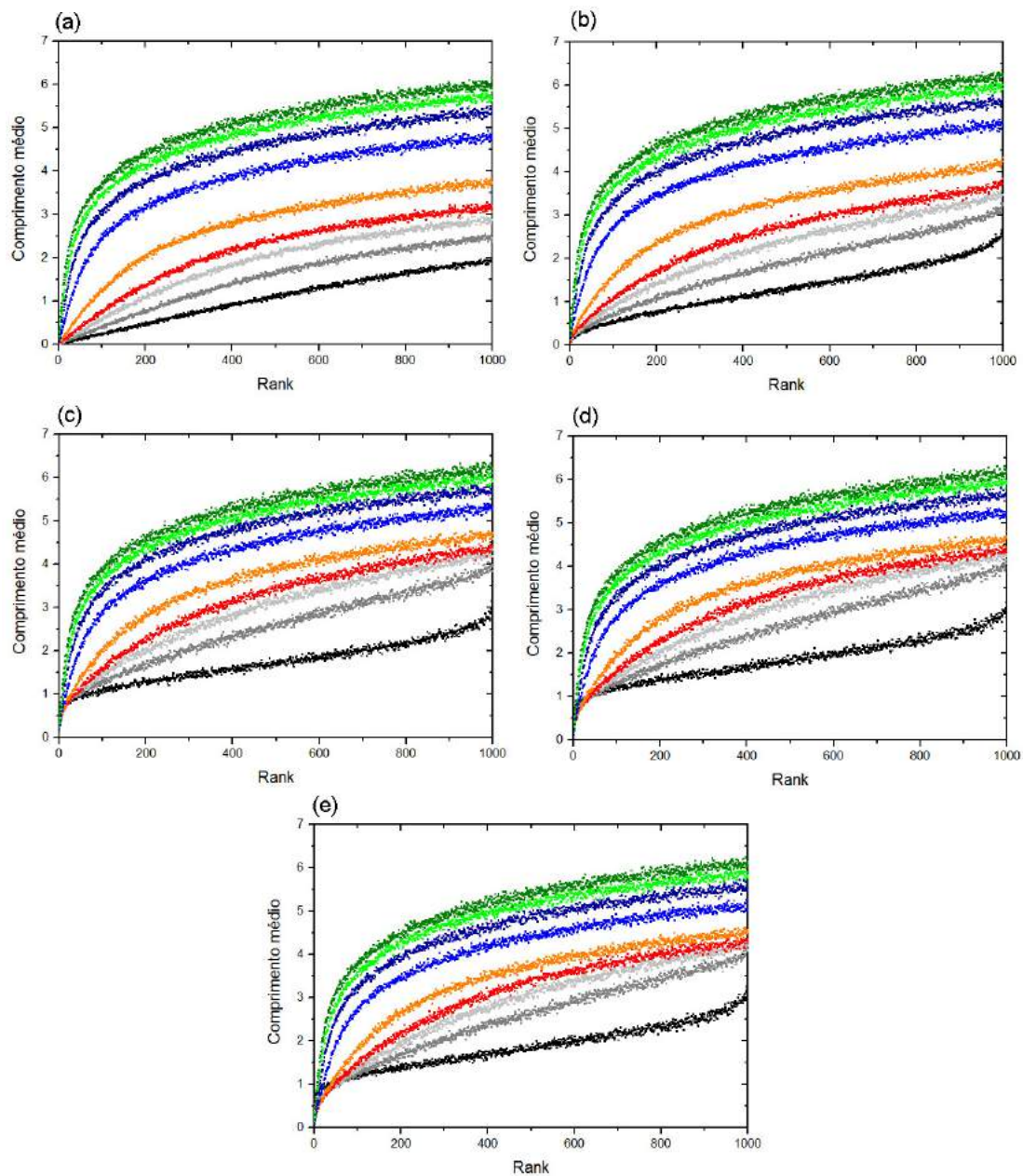


Figura 4.19: Rank das caminhadas padrão nos grafos aleatórios, com distribuição normal de fitness. Para as figuras (a), (b), (c), (d) e (e) temos $\lambda = 0, 0.5, 0.9, 0.95$ e 0.99 , respectivamente. Em cada gráfico, para cada cor corresponde um valor de z , a saber: $z = 2$ (preto), $z = 4$ (cinza escuro), $z = 6$ (cinza claro), $z = 8$ (vermelho), $z = 14$ (laranja), $z = 40$ (azul claro), $z = 72$ (azul escuro), $z = 104$ (verde claro), $z = 136$ (verde escuro).

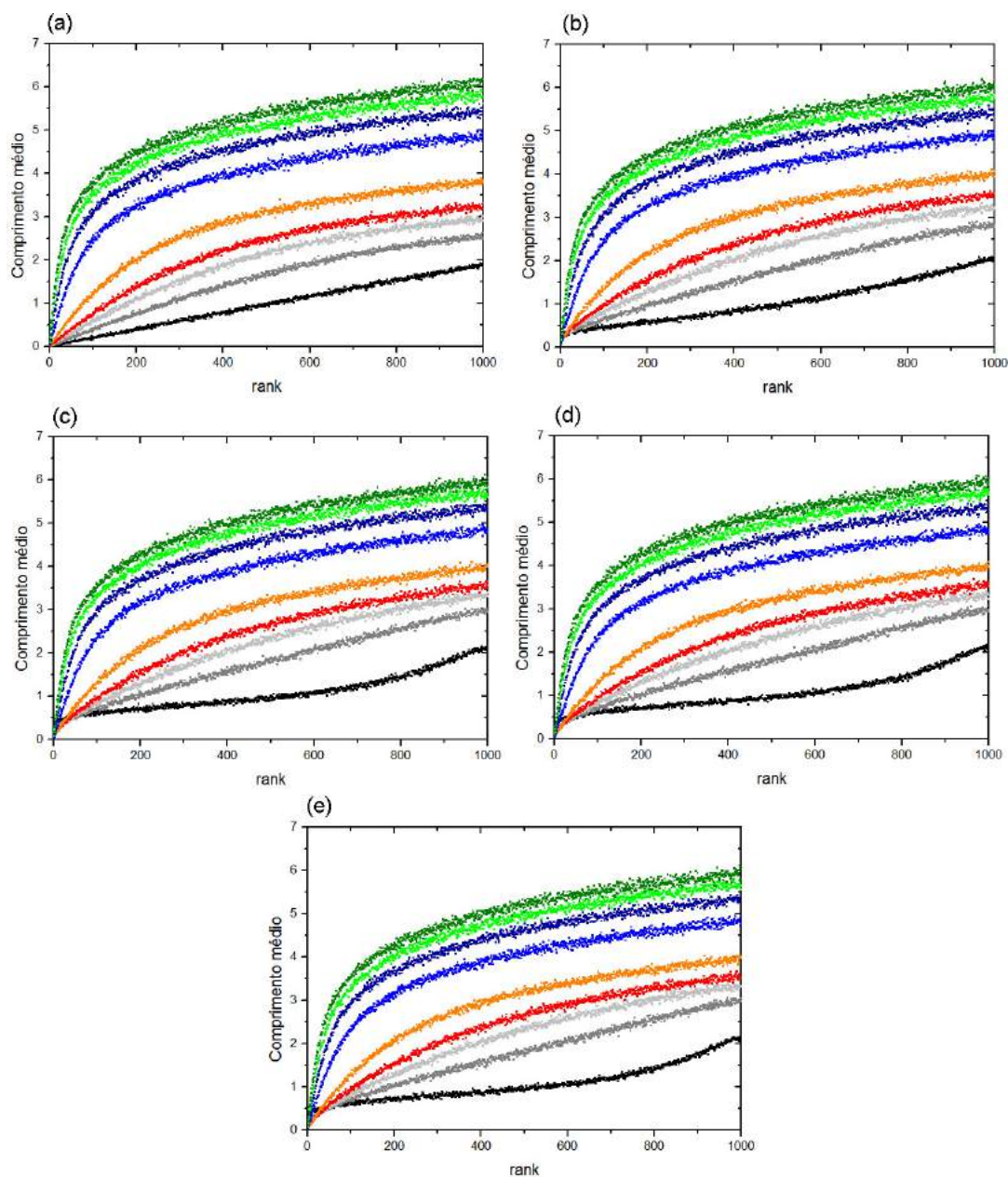


Figura 4.20: Rank das caminhadas padrão nas redes livres de escala, com distribuição de fitness exponencial bivariada. Para as figuras (a), (b), (c), (d) e (e) temos $\theta = 0, 0.5, 0.9, 0.95$ e 0.99 , respectivamente. Em cada gráfico, para cada cor corresponde um valor de z , a saber: $z = 2$ (preto), $z = 4$ (cinza escuro), $z = 6$ (cinza claro), $z = 8$ (vermelho), $z = 14$ (laranja), $z = 40$ (azul claro), $z = 72$ (azul escuro), $z = 104$ (verde claro), $z = 136$ (verde escuro).

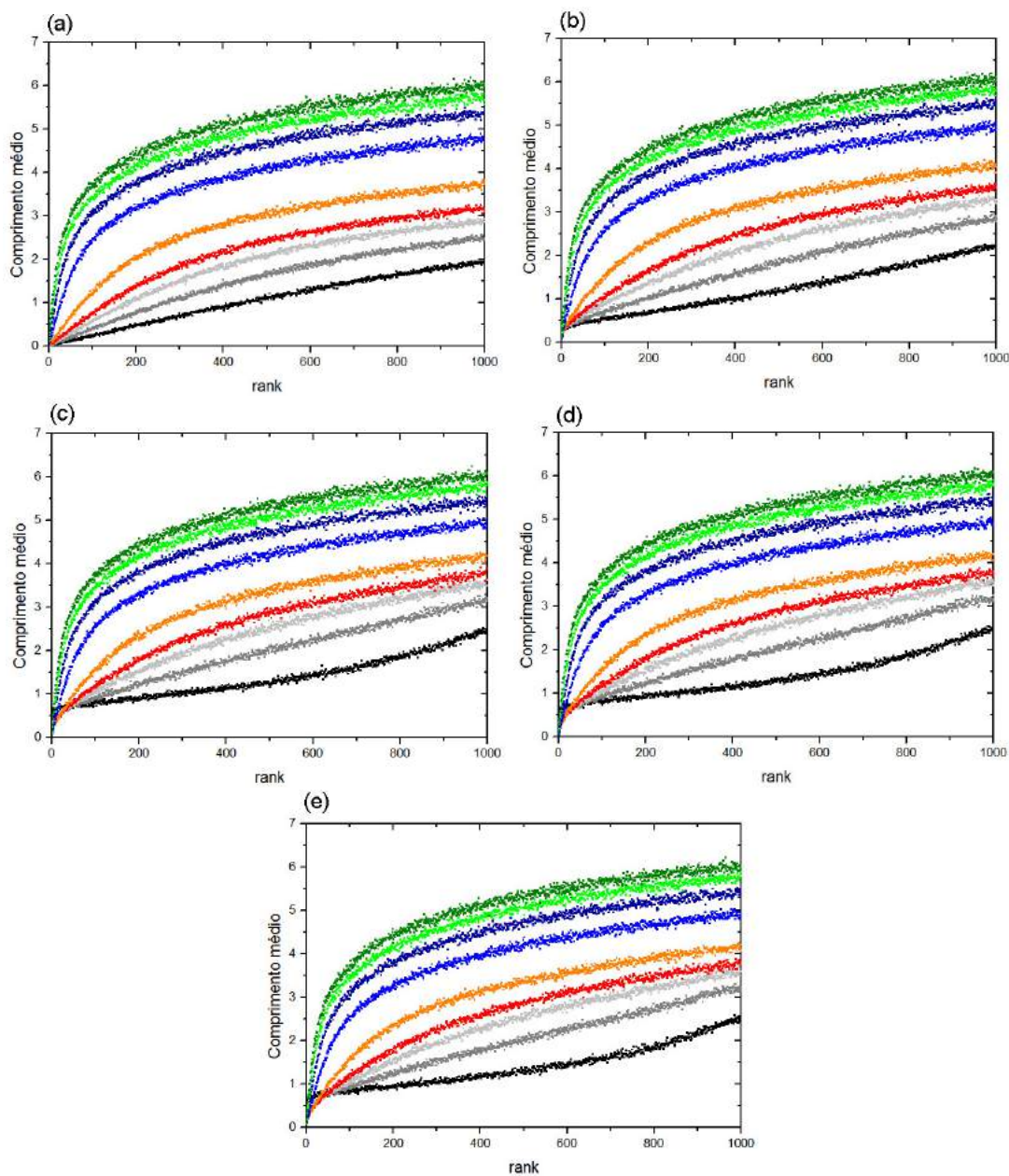


Figura 4.21: Rank das caminhadas padrão nos grafos aleatórios, com distribuição de fitness exponencial bivariada. Para as figuras (a), (b), (c), (d) e (e) temos $\theta = 0, 0.5, 0.9, 0.95$ e 0.99 , respectivamente. Em cada gráfico, para cada cor corresponde um valor de z , a saber: $z = 2$ (preto), $z = 4$ (cinza escuro), $z = 6$ (cinza claro), $z = 8$ (vermelho), $z = 14$ (laranja), $z = 40$ (azul claro), $z = 72$ (azul escuro), $z = 104$ (verde claro), $z = 136$ (verde escuro).

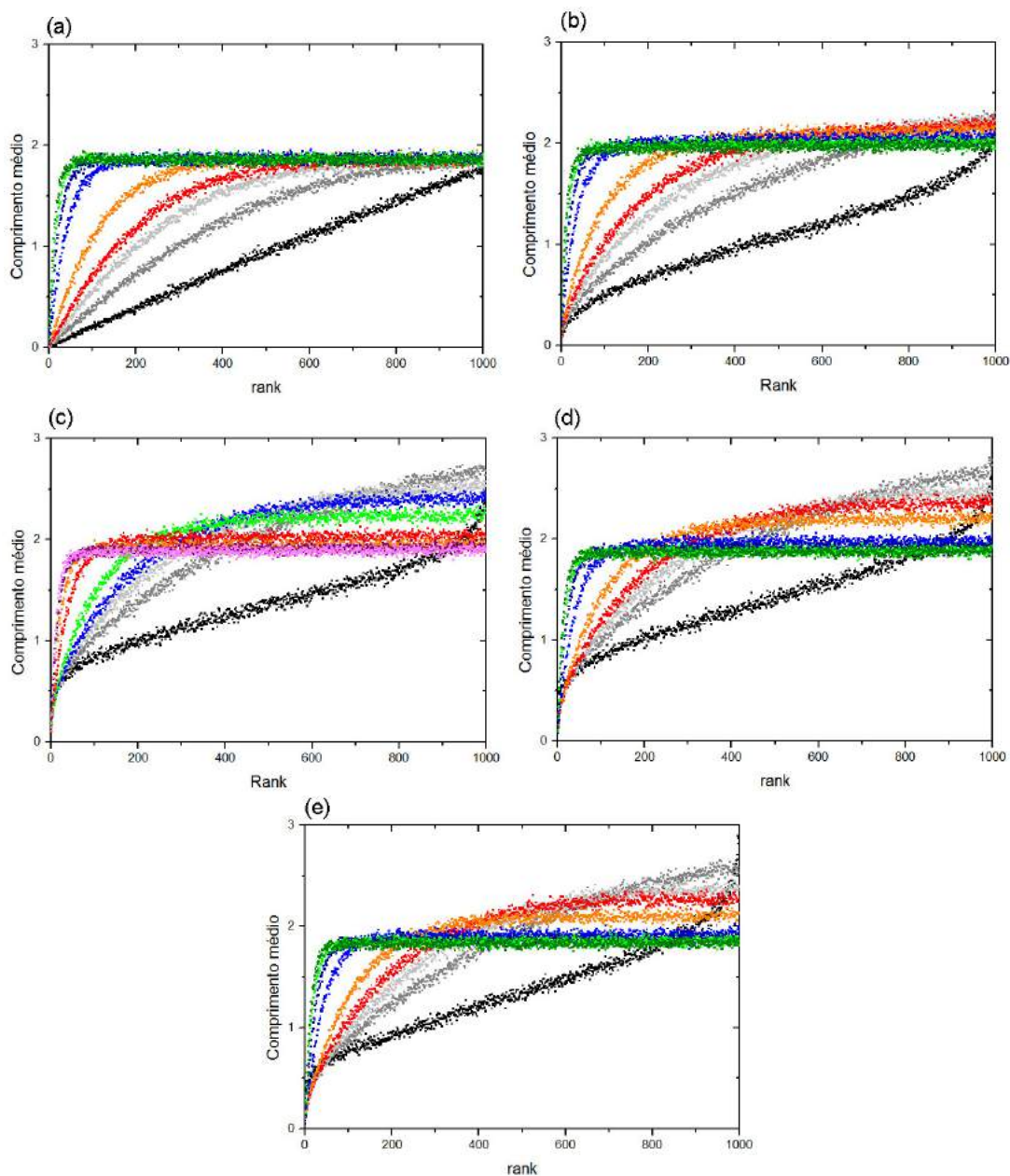


Figura 4.22: Rank das caminhadas gradiente nas redes livres de escala, com distribuição normal de fitness. Para as figuras (a), (b), (c), (d) e (e) temos $\lambda = 0, 0.5, 0.9, 0.95$ e 0.99 , respectivamente. Em cada gráfico, para cada cor corresponde um valor de z , a saber: $z = 2$ (preto), $z = 4$ (cinza escuro), $z = 6$ (cinza claro), $z = 8$ (vermelho), $z = 14$ (laranja), $z = 40$ (azul claro), $z = 72$ (azul escuro), $z = 104$ (verde claro), $z = 136$ (verde escuro).

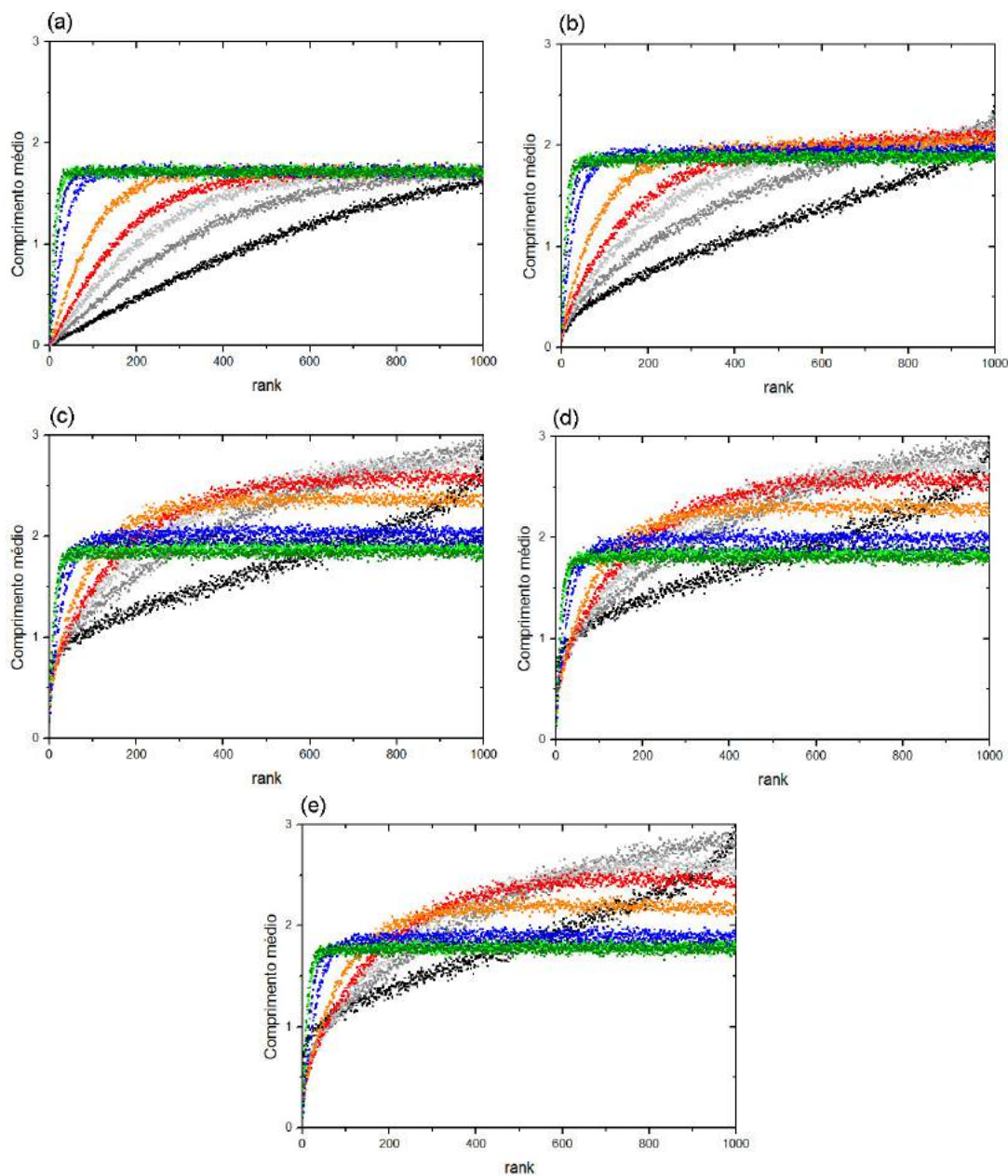


Figura 4.23: Rank das caminhadas gradiente nos grafos aleatórios, com distribuição normal de fitness. Para as figuras (a), (b), (c), (d) e (e) temos $\lambda = 0, 0.5, 0.9, 0.95$ e 0.99 , respectivamente. Em cada gráfico, para cada cor corresponde um valor de z , a saber: $z = 2$ (preto), $z = 4$ (cinza escuro), $z = 6$ (cinza claro), $z = 8$ (vermelho), $z = 14$ (laranja), $z = 40$ (azul claro), $z = 72$ (azul escuro), $z = 104$ (verde claro), $z = 136$ (verde escuro).

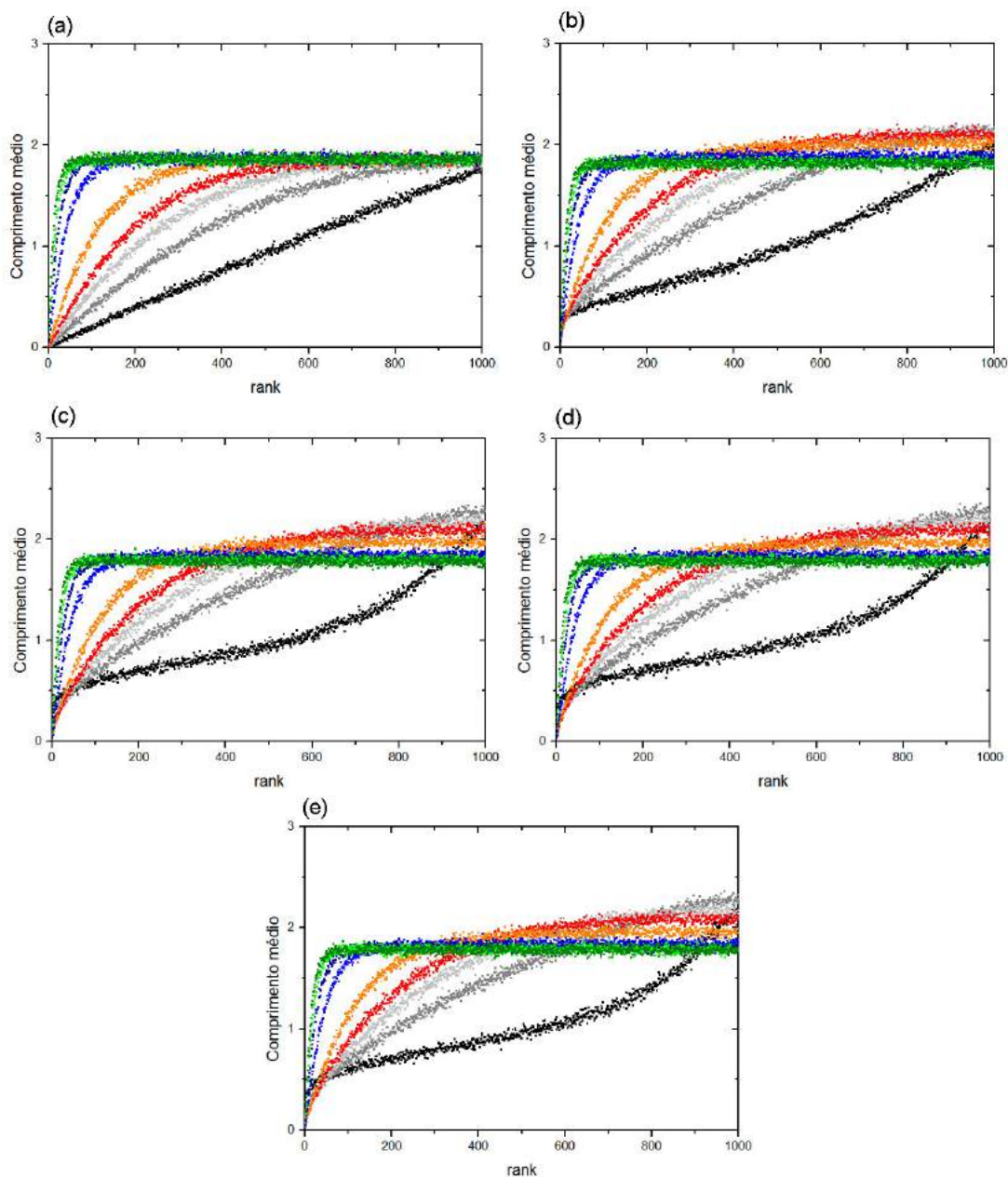


Figura 4.24: Rank das caminhadas gradiente nas redes livres de escala, com distribuição de fitness exponencial bivariada. Para as figuras (a), (b), (c), (d) e (e) temos $\theta = 0, 0.5, 0.9, 0.95$ e 0.99 , respectivamente. Em cada gráfico, para cada cor corresponde um valor de z , a saber: $z = 2$ (preto), $z = 4$ (cinza escuro), $z = 6$ (cinza claro), $z = 8$ (vermelho), $z = 14$ (laranja), $z = 40$ (azul claro), $z = 72$ (azul escuro), $z = 104$ (verde claro), $z = 136$ (verde escuro).

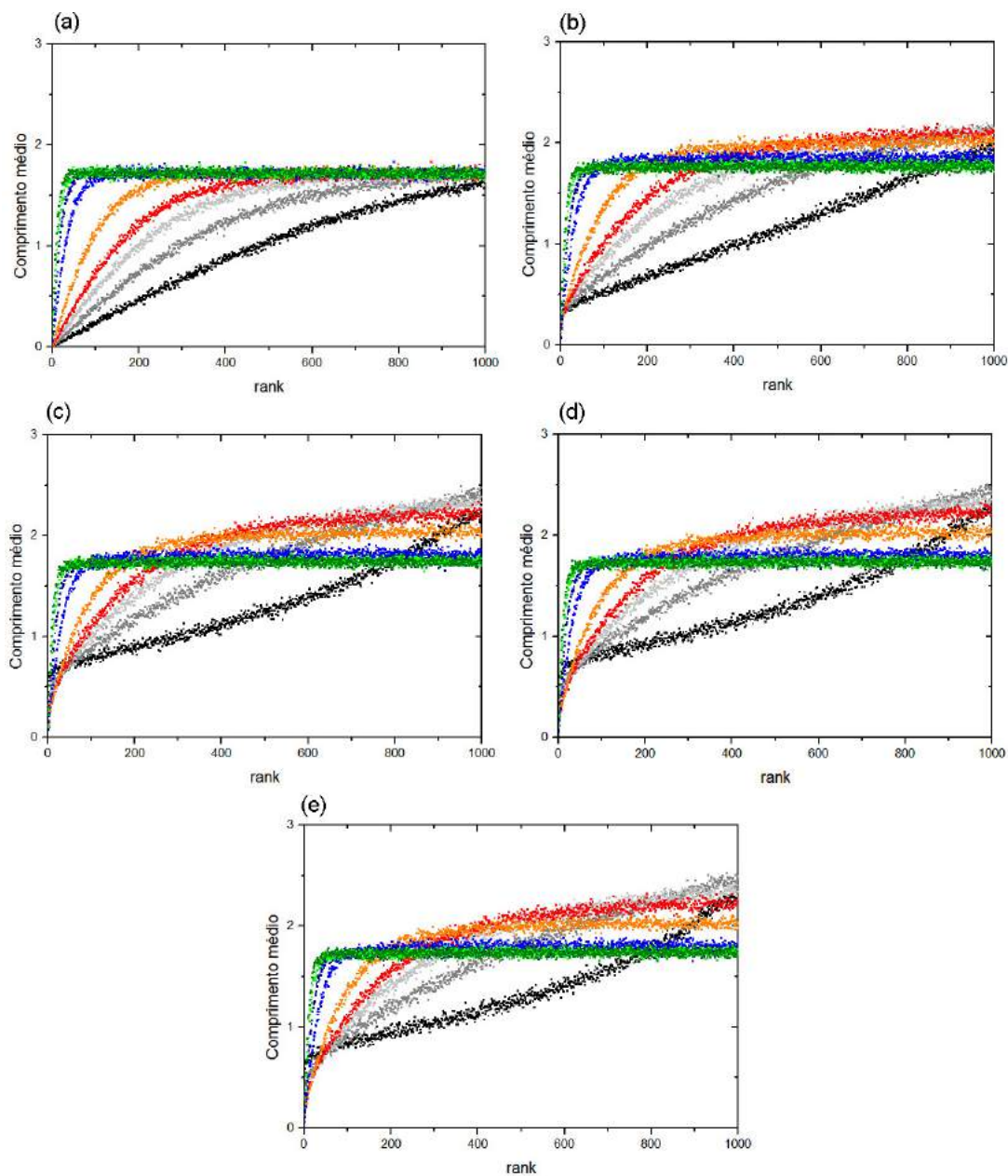


Figura 4.25: Rank das caminhadas gradiente nos grafos aleatórios, com distribuição de fitness exponencial bivariada. Para as figuras (a), (b), (c), (d) e (e) temos $\theta = 0, 0.5, 0.9, 0.95$ e 0.99 , respectivamente. Em cada gráfico, para cada cor corresponde um valor de z , a saber: $z = 2$ (preto), $z = 4$ (cinza escuro), $z = 6$ (cinza claro), $z = 8$ (vermelho), $z = 14$ (laranja), $z = 40$ (azul claro), $z = 72$ (azul escuro), $z = 104$ (verde claro), $z = 136$ (verde escuro).

5 CONCLUSÕES

As caminhadas adaptativas têm ocupado posição relevante nas pesquisas sobre biologia evolutiva. Nesse trabalho fizemos um estudo sobre as propriedades estatísticas de caminhadas adaptativas sobre relevos de adaptação correlacionados. Foram efetuadas caminhadas sobre redes livres de escala e sobre grafos aleatórios, associando a cada nó de cada rede um valor adaptativo (fitness) através das distribuições normal e exponencial bivariada. Esses valores adaptativos foram atribuídos de forma correlacionada, ou seja, o fitness de um nó está correlacionando com os fitness dos seus vizinhos.

Para cada tipo de rede complexa, com cada distribuição de fitness e para cada conectividade, estudamos o comprimento médio das caminhadas, a densidade média de ótimos locais, a diferença relativa entre os fitness local e global, a distribuição de frequências dos comprimentos das caminhadas e o rank das caminhadas. Foram feitas caminhadas dos tipos padrão e gradiente.

Quanto ao comprimento médio das caminhadas das caminhadas padrão, \bar{L} , como função da conectividade, observamos que não houve diferenças significativas em relação às diferentes topologias consideradas, bem como em relação às diferentes distribuições de fitness. Também não houve apreciáveis variações nos resultados quando alteramos o parâmetro de correlação. Para todos os casos de caminhada padrão, percebemos que o comprimento médio cresce com a conectividade z , com uma taxa de crescimento maior para pequenos valores de z e diminuindo essa taxa de crescimento para valores maiores de z . Em relação às caminhadas gradiente, observamos que para o relevo não-correlacionado, o comportamento qualitativo foi análogo ao obtido nas caminhadas padrão, independentemente da topologia ou da distribuição do fitness. Observamos que o número médio de passos é menor do que o observado nas caminhadas padrão. Para relevos correlacionados, verificamos um crescimento mais acentuado de \bar{L} para pequenos valores de z , atingindo um pico para um valor de z também pequeno, a partir daí decrescendo e assumindo um valor próximo do obtido para relevo não correlacionado, para valores maiores de z . O comportamento qualitativo não mudou para diferentes topologias e distribuições

do fitness.

Com relação à densidade média de ótimos locais, $\bar{\eta}$, em função da conectividade z , observamos um mesmo comportamento para as duas topologias consideradas, bem como para as diferentes distribuições de fitness. Em todos os casos observamos que o valor de $\bar{\eta}$ diminuiu com z de acordo com uma lei de potência $\bar{\eta} \propto z^{-\alpha}$, com α próximo de 0.9.

Com respeito à diferença relativa entre os valores adaptativos dos mínimos locais e global, $\overline{\Delta F_G}$, observamos que $\overline{\Delta F_G}$ decresce com z de acordo com uma lei de potência $\overline{\Delta F_G} \propto z^{-\beta}$, com β próximo de 0.2, independentemente da topologia considerada ou da distribuição do fitness. Verificamos que $\overline{\Delta F_G}$ diminui com a correlação quando a distribuição normal é considerada. Um comportamento oposto é observado para a distribuição exponencial. Num gráfico comparativo entre as caminhadas padrão e gradiente, constatamos que a caminhada gradiente é mais eficiente do que a caminhada padrão.

Com relação à distribuição de frequência dos comprimentos das caminhadas, não observamos diferenças importantes para as diferentes topologias ou para as diferentes distribuições de fitness. Foram observadas diferenças qualitativas apenas para os dois tipos diferentes de caminhadas realizadas (padrão e gradiente). Também não foram percebidas diferenças significativas para os diferentes parâmetros de correlação. Nas caminhadas padrão observamos que para valores pequenos de z temos uma maior frequência de caminhadas com comprimentos pequenos. Para conectividades grandes, observamos um aumento no tamanho médio do comprimento das caminhadas. Nas caminhadas gradiente, observamos uma predominância de caminhadas de comprimentos pequenos.

Quanto ao rank das caminhadas padrão, não percebemos diferenças qualitativas importantes para as diferentes topologias e distribuições de fitness consideradas. Também não observamos diferenças para os diferentes valores de λ . Para todos esses valores de λ observamos que quanto maior a conectividade z , mais rapidamente cresce o valor de \bar{L} . Para a caminhada gradiente, observamos comportamentos diferentes para diferentes valores de λ , apesar de não percebermos diferenças quando modificamos a topologia ou a distribuição do fitness.

Entendemos que nosso trabalho trará um estímulo para novos estudos sobre as caminhadas adaptativas sobre relevos de adaptação correlacionados, visto que poucos estudos foram feitos sobre caminhadas sobre relevos correlacionados.

Referências

- Albert, R. e Barabási, A. -L. (2002). Statistical Mechanics of Complex Networks. *Rev. Mod. Phys.* **74**, 47.
- Campos, P. R. A. e Moreira, F. G. B. (2005). Adaptive walk on complex networks. *Phys. Rev. E* **71**, 061921.
- Darwin, C. R. (1859). *The Origin of Species*. (J. Murray, London).
- de Oliveira, V. M. e Campos, P. R. A. (2005). Topological effects of fitness landscapes on the process of fixation of advantageous mutations. *Physica A* **346**, 451-458.
- Elena, S. F. e Lenski, R. E. (1997). Test of synergistic interactions among deleterious mutations in bacteria. *Nature (London)* **390**, 395.
- Erdős, P. e Rényi, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.* **5**, 17.
- Fisher, R. A. (1930). *The Genetical Theory of Natural Selection*. (Oxford Univ. Press, Oxford).
- Fontana, W., Konings, D. A., Stadler, P. F. e Schuster, P. F. (1993). Statistics of RNA Secondary Structures. *Biopolymers* **33**, 1389.
- Fontana, W., Schnabl, W. e Schuster, P. (1989). Physical aspects of evolutionary optimization and adaptation. *Phys. Rev. A* **40**, 3301.
- Gumbel, E. J. (1960). Bivariate exponential distributions. *Journal of the American Statistical Association* **55**, 698-707.
- Kauffman, S. (1993) *The Origins of Order*(Oxford University Press, Oxford).
- Kauffman, S. e Levin, S. (1987). Towards a General Theory of Adaptive Walks on Rugged Landscapes. *J. Theor. Biol.* **128**, 11.
- Lenski, R. E., Ofria, C., Collier, T. C. e Adami, C. (1999). Genome complexity, robustness and genetic interactions in digital organisms. *Nature (London)* **400**, 661-664.
- Lenski, R. E. e Travisano, M. (1994). Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. *Proc. Natl. Acad. Sci. USA* **91**, 6808.
- Maynard Smith, J. (1962). *The Scientist Speculates: an Anthology of Partly-Baked Ideas* (ed. Good, I. J.), 252256 (Basic Books, New York).

- Newman, M. E. (2001). The structure and function of complex networks. *Proc. Natl. Acad. Sci. USA* **98**, 404.
- Orr, H. A. (2003). A Minimum on the Mean Number of Steps Taken in Adaptive Walks. *J. Theor. Biol.* **220**, 241-247.
- Orr, H. A. (2005). The genetic theory of adaptation: a brief history. *Nature Reviews Genetics* **6**, 119-127.
- Rozen et al. (2002). Fitness Effects of Fixed Beneficial Mutations in Microbial Populations. *Current Biology* **12**, 1040 - 1045.
- Sanjuán, R., Moya, A. e Elena, S. F. (2004). The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc. Natl. Acad. Sci. USA* **101**, 8396.
- Wilke, C. O., Campos, P. R. A. e Fontanari, J. F. (2002). Genealogical process on a correlated fitness landscape. *J. Exp. Zool.: Mol. Dev. Evol.* **294**, 274.
- Wright, S. (1931). Evolution in mendelian populations. *Genetics* **16**, 97.