

DANIEL CARLOS DE OLIVEIRA

PROPAGAÇÃO DE DANOS EM UM MODELO DE REGULAÇÃO
GÊNICA EM REDES LIVRES DE ESCALAS

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Física Aplicada, para obtenção do título de *Magister Scientiae*.

VIÇOSA
MINAS GERAIS - BRASIL
2006

**Ficha catalográfica preparada pela Seção de Catalogação e
Classificação da Biblioteca Central da UFV**

T

O48p
2006

Oliveira, Daniel Carlos de, 1979-

Propagação de danos em um modelo de regulação
gênica em redes livres de escalas / Daniel Carlos de
Oliveira. - Viçosa, MG, 2006.

xv, 66f. : il. ; 29cm.

Orientador: Marcelo Lobato Martins.

Dissertação (mestrado) - Universidade Federal de
Viçosa.

Referências bibliográficas: f. 62-66.

1. Física estatística. 2. Automato celular. 3. Redes de
informação. 4. Genes. 5. Regulação de expressão gênica.

I. Universidade Federal de Viçosa. II. Título.

CDD 22.ed. 530.13

DANIEL CARLOS DE OLIVEIRA

PROPAGAÇÃO DE DANOS EM UM MODELO DE REGULAÇÃO GÊNICA EM
REDES LIVRES DE ESCALAS

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Física Aplicada, para obtenção do título de *Magister Scientiae*.

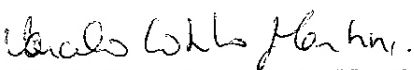
APROVADA: 15 de dezembro de 2006


Prof. Silvio da Costa Ferreira Junior
(Co-Orientador)


Prof. Marcelo José Vilela
(Co-Orientador)


Prof. Alcides Volpato Carneiro de Castro e Silva


Prof. Sidiney Geraldo Alves


Prof. Marcelo Lobato Martins
(Orientador)

Para Joel e Jussara,
meus pais queridos

AGRADECIMENTOS

- Quem almeja “passar além do Bojador tem que passar além da dor” e ao longo desse caminho muitas coisas aconteceram e muitas pessoas contribuíram para essa caminhada. Alguns estão com você do início ao fim da jornada. Outras conhecemos ao longo do caminho, mas cada um tem o seu próprio caminho a seguir, e nos separamos. Mas, todas foram importantes para o sucesso da viagem. Com certeza vou cometer dois erros: deixar de citar algumas pessoas e essas simples palavras não serão suficientes para expressar a minha imensa gratidão.
- Agradeço a Deus pela força durante todos esses dias.
- Aos meus queridos pais que sempre estiveram e estarão comigo.
- A todos os meus familiares que me ajudaram nessa caminhada.
- A querida Evelyn, obrigado por fazer parte da minha vida.
- Ao Prof. Marcelo Lobato Martins pela orientação.
- A todos professores e funcionários do departamento de Física UFV.
- Aos amigos verdadeiros que por sinal são muitos.

- Amigos de início da caminhada: Pablo, Léo, Marcelo, Édio, Flávia, Rafael, Paula, Paulo.
- Amigos da Física: Leandro, Kenedy, Jeferson, Beth, Fagner, Júlio, Tati, Carlos, Jackson, Allison, Samir, Igor, Thiago, Paulo Henrique, Fábio (Carioca), Ney (Gaúcho), Jaqueline, Sandro, Luciano, Fernanda e outros que com certeza esqueci.
- Amigos de República: Leandro, Maurício, Carlos, Júlio, Dinho, Jean, Matheus, Bruno, Felipe, Lucas, Mycon, Mateus, Marcel e outros.
- A Universidade Federal de Viçosa - UFV.
- A Capes pelo apoio financeiro.

BIOGRAFIA

DANIEL CARLOS DE OLIVEIRA, filho de Jussara Maria de Oliveira e Joel Carlos de Oliveira, nasceu em Governador Valadares, Minas Gerais, em 03 de agosto de 1979. Em julho de 2004, concluiu a licenciatura e o bacharelado em Física pela Universidade Federal de Viçosa. Em dezembro de 2006, obteve o título de Mestre em Física Aplicada na Universidade Federal de Viçosa.

SUMÁRIO

LISTA DE FIGURAS	viii
LISTA DE TABELAS	xi
RESUMO	xii
ABSTRACT	xiv
1 Introdução	1
2 O Problema Biológico	3
2.1 O Controle de Genes	3
2.2 Redes de Regulação Gênica	7
2.3 Redes Complexas	10
2.4 As redes biológicas e o modelo de Barabási-Albert	21
3 Modelos de Controle de Genes e de Diferenciação Celular.	26
3.1 Autômatos Celulares	26
3.2 O modelo de Kauffman	28
3.3 O modelo de Sales-Martins-Stariolo (SMS)	31
3.4 Resultados básicos destes modelos	34
4 O Modelo de Sales-Martins-Stariolo em Redes Livres de Escalas	40
4.1 Estrutura de atratores	40
4.2 A propagação de danos	41

4.3 Resultados e Discussão	44
5 Conclusões	58
6 Perspectivas Futuras	60
Referências Bibliográficas	62

LISTA DE FIGURAS

2.1	Descrição esquemática do DNA.	5
2.2	Estrutura hierárquica da organização molecular em uma célula.	6
2.3	Molécula de DNA durante o processo de replicação, passando pela transcrição do RNA e a tradução no produto final, a proteína.	8
2.4	O dogma central da biologia molecular com o fluxo de informação fluindo através da replicação (DNA), da transcrição (RNA) e tradução (Proteína).	9
2.5	Proteína regulatória que ativa ou desativa conjuntos específicos de genes.	10
2.6	Esquemas de redes com seus sítios e arestas. (a) Uma rede de presas-predadores. (b) Rede de colaborações entre cientistas. (c) Rede de contatos sexuais entre seres humanos.	11
2.7	(a) Grafos aleatórios. (b) Distribuição do número de conectividades de cada sítio.	12
2.8	Modelo WS que inclui desde as redes regulares ($p = 0$) até redes aleatórias ($p = 1$).	15
2.9	Rede livre de escala com um pequeno número de sítios e arestas.	16
2.10	(a) Rede de interação de proteínas da levedura <i>Saccharomyces cerevisiae</i> . (b) Distribuição de conectividade dos sítios.	17

2.11	Esquema de uma rede com topologia livre de escala, para o caso de $\gamma = 2.5$	20
3.1	Padrões espaço-temporais das classes de Wolfram.	29
3.2	Diagrama de fases do Modelo de Kauffman.	35
3.3	Diagrama de fases do Modelo de Sales-Martins-Stariolo	39
4.1	Esquema de partição do espaço de estados por seus atratores.	42
4.2	Distância de Hamming em função do tempo.	45
4.3	Dano final médio em função do dano inicial para diferentes valores m de ligações iniciais por gene e $\theta = 0$	47
4.4	Dano final médio em função do dano inicial para $\theta = 0.8$	48
4.5	Dano final médio em função de θ	49
4.6	Diagrama de fases do modelo Sales-Martins-Stariolo na rede livre de escala de Albert-Barabási.	51
4.7	Gráfico da distribuição de conectividades dos sítios de uma rede genética com $N = 80000$ genes.	52
4.8	Dano final médio em função de θ para alguns valores de γ	54
4.9	Distribuição de conectividades para valores diferentes de γ	55
4.10	Diagrama de fases para o modelo de Sales-Martins-Stariolo para redes genéticas com topologia livre de escala.	55
4.11	Período dos atratores no modelo de Sales-Martins-Stariolo em função do número de genes N na rede genética.	56

4.12	Número médio de atratores em função do tamanho da rede genética para valores de γ diferentes, proposto por Aldana	56
4.13	Número médio de atratores (células diferentes) no modelo de Sales-Martins-Stariolo como função do número de genes N	57

LISTA DE TABELAS

2.1	Características gerais de algumas redes reais comparadas com modelos de redes aleatórias	21
2.2	Parâmetros de redes reais comparados com redes com topologia livre de escala	21

RESUMO

OLIVEIRA, Daniel Carlos de, M.Sc., Universidade Federal de Viçosa, dezembro de 2006. **Propagação de Danos em um Modelo de Regulação Gênica em Redes Livres de Escalas.** Orientador: Marcelo Lobato Martins. Co-Orientadores: Sílvio da Costa Ferreira Júnior e Marcelo José Vilela.

A compreensão de todos os elementos funcionais que controlam a expressão gênica é um desafio da biologia molecular moderna. Vários avanços no entendimento do funcionamento das redes de genes foram feitos com a descrição desse sistema complexo por meio de modelos de redes. A caracterização estrutural dessas redes permite conhecer como tais sistemas são organizados e permite construir modelos quantitativos para a dinâmica de interação que envolve um grande número de genes e uma elevada complexidade. O ponto central do nosso trabalho é a análise da estabilidade frente a mutações de uma rede de interação gênica descrita por um modelo de autômatos celulares proposto por Sales-Martins-Stariolo em redes livres de escalas. Redes livres de escalas são caracterizadas por distribuições de graus de conectividades dos sítios (genes) descritas por leis de potência. A dinâmica deste modelo apresentou um comportamento caracterizado por duas fases distintas: congelada e caótica. O número de tipos celulares diferentes (atratores) parece crescer exponencialmente com o número N de genes na rede na fase caótica e linearmente na fase congelada. Na região congelada a maioria dos genes está no estado ativo ou inativo, os períodos dos

atratores são pequenos, todos com praticamente o mesmo valor, crescendo linearmente com N e a rede é estável a mutações. Na região caótica temos atratores com períodos que crescem mais rapidamente, ainda de forma linear, e mesmo um pequeno dano ou mutação introduzido na rede se propaga para uma fração finita dos sítios. O sistema é, portanto, extremamente sensível às condições iniciais e instável. As fronteiras entre as fases congelada e caótica no espaço de parâmetro do modelo de Sales-Martins-Stariolo para regulação gênica em redes livres de escalas foram determinadas usando simulações computacionais.

ABSTRACT

OLIVEIRA, Daniel Carlos de, M.Sc., Universidade Federal de Viçosa, December, 2006. **Damage Spreading in a Model of Gene Regulation in Scale-Free Networks.** Adviser: Marcelo Lobato Martins. Co-Advisers: Sílvio da Costa Ferreira Júnior and Marcelo José Vilela.

The understanding of all functional elements that control genetic expression is a challenge of modern molecular biology. Several advances towards this goal have been made by describing this complex system through network models. The structural characterization of gene networks leads to fundamental insights about the systems' organization and provides the bases to build quantitative models for the dynamics of the interactions, involving a great number of genes. The central aim of our work was analysis the stability to mutations of a cellular automata model for gene regulation proposed by Sales-Martins-Stariolo in scale-free networks. We find that the dynamics of this model exhibits a behavior characterized by two distinct phases: frozen and chaotic. The number of differentiated cell types (attractors) increases exponentially in the chaotic phase and in the frozen phase scales linearly with the number N of genes in the network. The majority of the genes in the frozen region are fixed in an active or inactive state. The periods of the attractors are small, all of them having practically the same length. This average period scales linearly with N and the network is stable

to mutations. In contrast, in the chaotic region even a minimal damage or mutation introduced in the network spreads throughout a finite fraction of nodes and periods of the attractors grow faster, but again in a linear way. Therefore, the system is extremely sensitive to the initial conditions. The edge between the frozen and chaotic phases in the parameter space of the Sales-Martins-Stariolo model for gene regulation in scale-free networks was determined by computer simulation.

Capítulo 1

Introdução

A imensa diversidade dos organismos vivos é fascinante e caracterizada por uma complexidade que permeia as mais variadas escalas, desde a constituição molecular, por exemplo, até as redes sociais de interações. A ciência tenta compreender as leis que regem toda essa complexidade. Uma chave para tal entendimento surge quando se observa que a partir de leis simples, estruturas complexas podem emergir. A Física Estatística, juntamente com a teoria de redes, contribuiu enormemente para o entendimento dos sistemas complexos. Um sistema complexo pode ser definido como sendo conjuntos de componentes relativamente simples que interagem entre si e mostram a capacidade de criar padrões espaço-temporais os mais variados possíveis. A ciência que estuda a complexidade relata que leis dinâmicas simples são suficientes para levar ao surgimento de padrões diversificados após repetidas iterações. A quebra de simetria, a instabilidade e a criticalidade auto-organizada são argumentos que permeiam esse ramo da ciência.

Uma das estratégias para se compreender essa diversidade de organismos vivos passa pela modelagem das redes de interações genéticas, constituintes da unidade

básica de formação dos seres vivos: a célula. As redes genéticas possuem em sua estrutura as informações necessárias para o estabelecimento da diversidade organizacional dos seres vivos. Os estudos desses sistemas complexos permitiram a criação de hipóteses acerca do funcionamento da estrutura celular. O conhecimento das partes que constituem as redes genéticas, os genes, não esclarece como os organismos vivos são formados. Entretanto, a dinâmica do sistema pode possibilitar a compreensão do funcionamento dos organismos vivos.

O objetivo da presente pesquisa foi a modelagem da rede de interação genética dos seres vivos. A compreensão do controle da expressão gênica que comanda a diferenciação celular permite o entendimento da diversidade celular.

A presente dissertação está organizada da seguinte forma: no capítulo 2 (O Problema Biológico) introduzimos uma breve revisão acerca da estrutura biológica que molda uma rede genética. Discutimos inicialmente a regulação gênica que ocorre no núcleo celular e fazemos uma revisão sobre a teoria de redes. Os modelos que tentam descrever a complexidade de uma rede de interação genética são tratados no capítulo 3 (Modelos de controle de genes e de diferenciação celular). Mais adiante, no capítulo 4 (O Modelo de Sales-Martins-Stariolo em Redes Livres de Escalas) apresentamos a teoria dos atratores, a propagação de danos e nossos principais resultados. Finalmente, no capítulo 5 estão sintetizados os principais resultados obtidos nessa dissertação e no capítulo 6 apresentamos as perspectivas futuras desse trabalho.

Capítulo 2

O Problema Biológico

Neste capítulo vamos apresentar as idéias básicas referentes às interações gênicas que são responsáveis pela diferenciação celular e uma introdução ao estudo das redes complexas.

2.1 O Controle de Genes

Um dos problemas centrais da biologia molecular moderna é o entendimento do conjunto de interações bioquímicas complexas que ocorrem nas células. Como exemplo, temos as interações entre genes mediadas pelas proteínas que eles codificam. O controle dessas interações, responsáveis pela expressão gênica, apresenta-se como um desafio para a ciência. Vencer este desafio pode levar à compreensão de como o metabolismo celular normal funciona e permitir o tratamento de suas alterações patológicas (terapias genéticas) [1]. Uma peça primordial dos mecanismos bioquímicos complexos nos organismos vivos é o gene. O gene é conceituado como a unidade fundamental de informação dos organismos vivos. Bioquimicamente, o gene é um segmento de DNA que codifica uma informação a qual será usada para a síntese de

um produto biológico funcional. Normalmente o produto gênico funcional é uma proteína.

Toda a informação genética dos organismos vivos está contida em uma macromolécula chamada ácido desoxirribonucleico (DNA). O DNA é formado por subunidades constituídas por uma base nitrogenada, um fosfato e uma pentose. Essas subunidades são chamadas nucleotídeos. Existem quatro bases nitrogenadas diferentes no DNA: adenina (A), timina (T), citosina (C) e guanina (G). Para formar uma molécula de DNA os nucleotídeos adjacentes são ligados entre si por interações fosfodiésteres. Além disso, cada base nitrogenada encontra-se ligada por pontes de hidrogênio a outra base específica, a saber: adenina com timina e guanina com citosina. Se pensarmos no material genético humano, temos entre 2,8 milhões e 3,5 milhões de pares desse tipo. Devido a essas interações, a molécula de DNA encontra-se na forma de dupla hélice. No DNA o açúcar é a desoxirribose. Na figura 2.1, temos um desenho esquemático do DNA.

Em um segmento de DNA, temos o gene que é responsável pela codificação de uma proteína. Os genes são agrupados em cromossomos. O número de cromossomos presentes na espécie humana é 46, 23 provenientes do pai e 23 da mãe. Esses 23 pares de cromossomos estão presentes no núcleo da célula. A figura (2.2) esquematiza toda essa rede complexa da organização molecular.

Células diferentes em um organismo multicelular diferem drasticamente em função e estrutura, mesmo com o genoma (conjunto de genes de um organismo) idêntico [2]. Para o caso de um organismo multicelular, temos várias células que se diferenciam umas das outras. A evidência concreta desse fato deve-se à síntese e ao

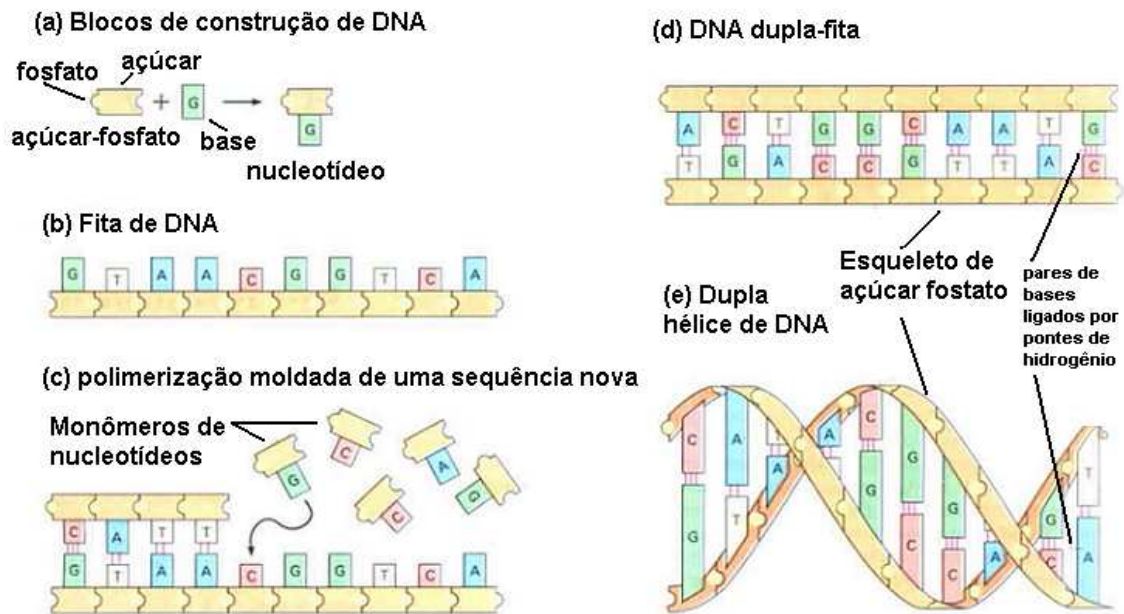


Figura 2.1: O DNA é formado por subunidades simples, chamadas nucleotídeos, constituídas de açúcar, fosfato e base nitrogenada. Uma cadeia simples de DNA consiste de nucleotídeos ligados por ligações fosfodiésteres entre açúcar e fosfato. As duas fitas formam uma dupla hélice. Figura adaptada da Ref. [2].

acúmulo de diferentes moléculas de proteínas nessas células. Uma etapa intermediária necessária à síntese protéica é a síntese de mRNA. O mRNA (ácido ribonucléico) é um polímero formado por um açúcar ribose, fosfato e por quatro bases: uracila (U), adenina (A), citosina (C) e guanina (G). Em cada célula são encontrados três tipos de RNAs: transportador, ribossômico e mensageiro. Os RNAs responsáveis pela codificação de aminoácidos de uma proteína são os RNA mensageiros (mRNAs). Análises dos genomas de células diferentes mostram que modificações na expressão de genes, base do desenvolvimento de organismos multicelulares [2], não são acompanhadas por mudanças do DNA. Ou seja, a diferenciação celular deve-se ao fato de existirem interações entre genes na rede genética, propiciando a formação de mRNAs de constituições diferentes.

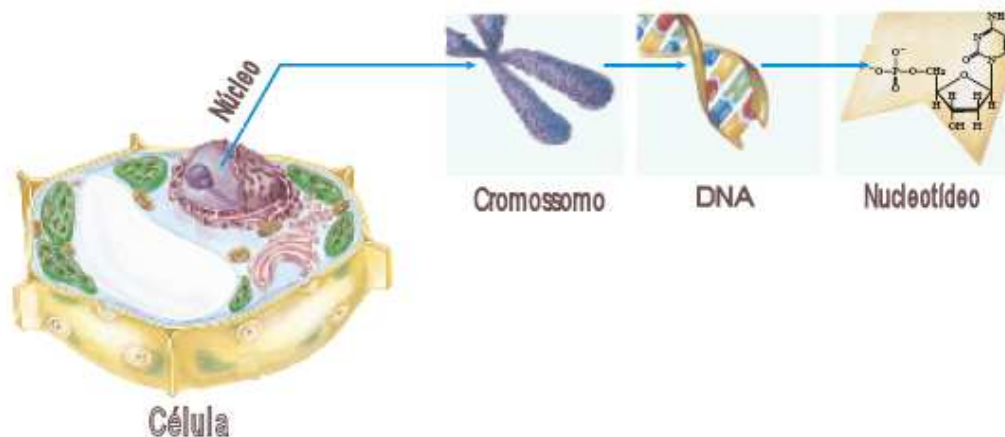


Figura 2.2: Estrutura hierárquica da organização molecular em uma célula. No núcleo temos os cromossomos que consistem de macromoléculas de DNA e de algumas proteínas. O DNA em forma de espiral com dupla hélice é formado de subunidades, os nucleotídeos. Figura adaptada da Ref. [3].

Células que se diferenciam umas das outras sintetizam conjuntos de proteínas diferentes. Um exemplo, as hemácias (glóbulos vermelhos do sangue) apresentam uma grande quantidade da proteína hemoglobina (responsável pelo transporte de oxigênio), ao passo que nenhuma das demais células do organismo apresentam a mesma proteína. Os fatores determinantes que controlam a quantidade em que cada proteína é sintetizada e mantida em uma célula são: a taxa de síntese de mRNAs correspondentes às proteínas, a frequência com que esses mRNAs são traduzidos e a própria estabilidade das proteínas produzidas. A taxa em que a síntese protéica ocorre depende do controle genético interno, e parcialmente do ambiente químico externo.

As células com funções diferenciadas sintetizam e mantêm diferentes conjuntos de moléculas e proteínas que são responsáveis pelas funções específicas de cada célula. A atividade genética diferenciada que ocorre devido às diversas interações protéicas é responsável pela diferenciação celular.

2.2 Redes de Regulação Gênica

Para entendermos uma rede de regulação gênica, devemos nos ater à estrutura do DNA. O DNA é uma molécula fundamental a todos os seres vivos e a maior macromolécula nas células. Um empacotamento de DNA e proteínas constitui um cromossomo e este pode carregar milhares de genes. Ele traz na sua seqüência, todas as informações genéticas de um indivíduo e possui a capacidade de transmiti-las aos seus descendentes. A molécula de DNA para desempenhar suas funções, deve ser capaz de sofrer os processos de replicação e transcrição. Por replicação, temos o processo de cópia do próprio DNA que é necessário para garantir a hereditariedade. Durante a replicação uma molécula de DNA origina duas moléculas idênticas. Já no processo de transcrição, ocorre a síntese de mRNA a partir do molde de DNA. Esse mRNA será posteriormente traduzido, originando dessa forma o produto gênico, ou seja, a proteína (Figura 2.3).

Na molécula de mRNA, existe um mecanismo para a propagação de informações contidas na molécula de DNA. Podemos pensar em um fluxo de informações que parte do DNA e é transmitido para o mRNA e por fim armazenado na proteína. Esse modelo unidimensional de informações [4] foi proposto pelo físico Francis Crick e é conhecido como dogma central da biologia molecular (Figura 2.4). Nele F. Crick relaciona DNA, RNA e proteínas. Esse fluxo é o meio pelo qual as células expressam suas informações gênicas.

O processo da transcrição acontece quando uma enzima específica, a RNA polimerase, liga-se a uma região do DNA conhecida como promotora. Essa região

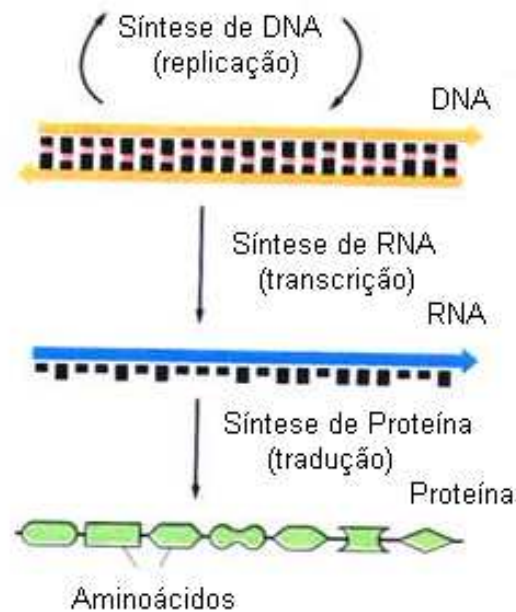


Figura 2.3: A seqüência que parte do DNA até o produto final a proteína. Na replicação temos uma cópia do DNA e na transcrição uma fita de DNA é usada como molde para a síntese de RNA. Logo em seguida, o RNA é usado para guiar a síntese da molécula de proteína. Figura retirada da Ref. [2].

orienta onde iniciar a síntese do mRNA. Quando a RNA polimerase termina a transcrição, ela libera o molde de DNA e a molécula de mRNA recém-sintetizada. A RNA polimerase reconhece apenas a região do DNA da seqüência promotora. Assegurando que somente esses genes serão transcritos.

Estudos feitos em bactérias nos anos 50 trouxeram evidências iniciais da existência de proteínas regulatórias de genes que ativam ou desativam conjuntos específicos de genes [2]. O DNA também possui seqüências regulatórias que são necessárias para ligar e desligar o gene. As seqüências regulatórias do DNA devem ser reconhecidas por proteínas específicas chamadas de proteínas regulatórias. As proteínas que reconhecem uma seqüência específica do DNA fazem isso porque sua superfície se ajusta às características espaciais da superfície do DNA. Como exemplo, vamos des-

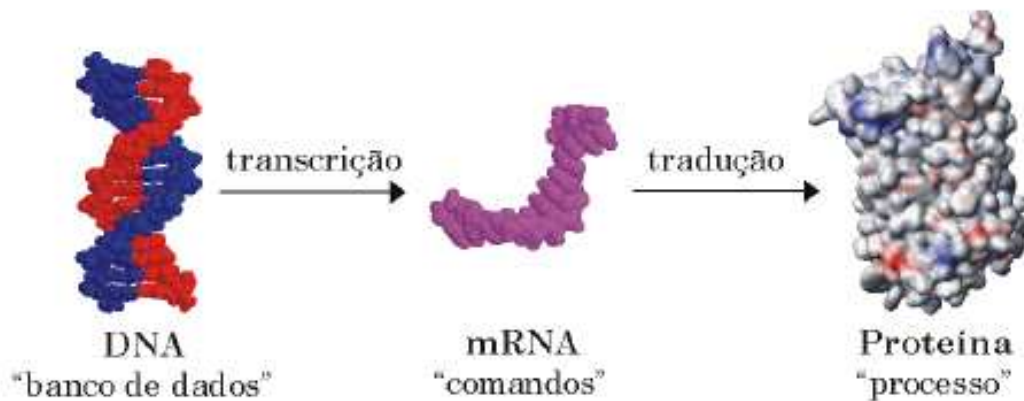


Figura 2.4: O dogma central da biologia molecular que traz o fluxo de informação fluindo através da replicação (DNA), da transcrição (RNA) e tradução (Proteína). Figura retirada da Ref. [4].

crever o processo que ocorre na bactéria *Escherichia coli*, um organismo unicelular que apresenta uma molécula de DNA circular de aproximadamente $4,6 \times 10^6$ pares de nucleotídeos. Este DNA codifica aproximadamente 4300 proteínas. A expressão de algumas proteínas é regulada de acordo com os nutrientes disponíveis no ambiente. Se o ambiente dessa bactéria contiver, por exemplo, o aminoácido triptofano, não há necessidade da síntese desse aminoácido. Então, uma proteína regulatória liga-se à cadeia de DNA que codifica as enzimas responsáveis pela síntese do triptofano na região chamada de operador impedindo o acesso da enzima RNA polimerase aos genes estruturais. Esse mecanismo impede a transcrição e por conseqüência a biossíntese de triptofano. Essa proteína regulatória é conhecida como repressor de triptofano. O repressor é um artifício que impede a síntese de proteínas específicas quando as mesmas não são necessárias à célula. Na figura 2.5 temos o desenho esquemático que apresenta o repressor.

Já nos anos 60 a biologia molecular teve um grande avanço na compreensão da regulação da atividade gênica com os pesquisadores François Jacob e Jacques

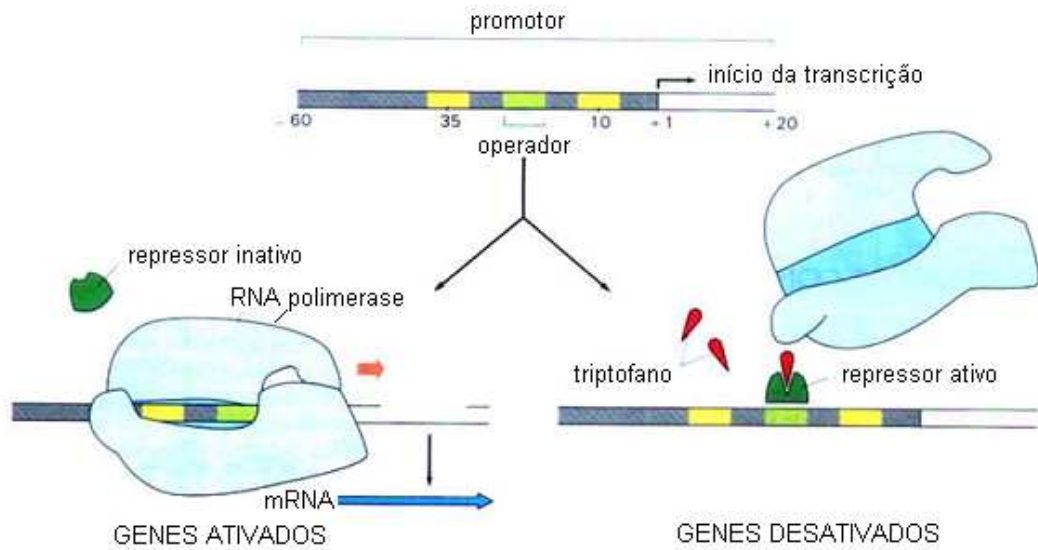


Figura 2.5: Se o nível de triptofano dentro da célula estiver baixo, a RNA polimerase liga-se ao promotor e transcreve o gene que codifica o triptofano. Se por outro lado existir uma grande quantidade de triptofano, o repressor do triptofano é ativado. Figura retirada da Ref. [2].

Monod [3]. Eles descobriram a existência de mecanismos de retroalimentação pelos quais os genes são ligados e desligados, análogo ao sistema binário dos computadores digitais. O modelo de regulação gênica, proposto nesse trabalho, leva esse fato em consideração. Os genes são tratados como elementos Booleanos capazes de assumir apenas dois estados: 1 (ativo) ou 0 (inativo), como veremos mais adiante.

2.3 Redes Complexas

Nos últimos anos diversos trabalhos foram publicados utilizando-se redes na modelagem de sistemas complexos. Qualquer sistema que possa ser dividido em elementos que se relacionam por regras podem ser tratados como redes. Redes são conjuntos de itens conectados entre si que denominamos vértices, nós ou sítios [5] onde as conexões são feitas por arestas ou ligações (Figura 2.6). Como exemplos de

redes encontradas na natureza ou geradas por processos tecnológicos, podemos citar: Internet, redes sociais, redes metabólicas, redes neurais, redes genéticas e inúmeras outras. A idéia de rede está presente desde o nível atômico até o mais complexo nível social observado na organização dos seres vivos. Portanto, o estudo de redes pode ser entendido como uma das chaves para a compreensão do universo que nos cerca.

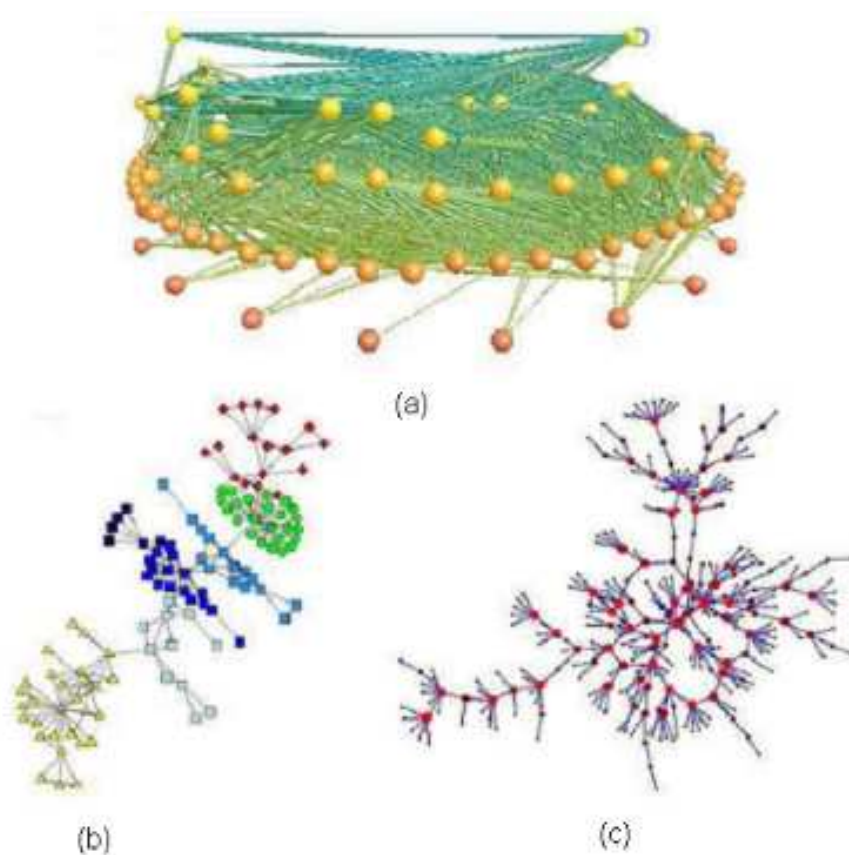


Figura 2.6: Esquemas de redes com seus sítios e arestas. (a) Uma rede de presas-predadores. (b) Rede de colaborações entre cientistas. (c) Rede de contatos sexuais entre seres humanos. Figura extraída da Ref. [5].

As primeiras idéias na teoria de redes foram propostas por Euler no século XVIII. Ele foi um dos pioneiros que ajudaram a fundamentar a Teoria dos Grafos. Matematicamente, um conjunto de sítios e arestas formam um grafo. Os sítios podem estar conectados ou não. As arestas servem para indicar uma relação existente entre

sítios conectados. Para cada modelo de rede proposto, temos um conjunto de relações entre os sítios. O estudo das redes por meio da Teoria dos Grafos é considerado um dos pilares da matemática discreta [5]. Os matemáticos húngaros Paul Erdős e Alfred Rényi fundamentaram o estudo dos grafos ao criarem uma teoria, também conhecida como modelo de Erdős-Rényi ou dos grafos aleatórios [6]. O modelo parte de um número n de sítios totalmente desconectados e em seguida conecta-se aleatoriamente cada par de sítios com probabilidade p . Os grafos gerados possuem $pN(N - 1)/2$ ligações distribuídas aleatoriamente. Se analisarmos a distribuição de conectividade dos sítios dessa rede, temos uma distribuição de Poisson. Essa distribuição informa como ocorrem as ligações dos sítios. Ou seja, a maioria dos sítios possui aproximadamente o mesmo número de ligações, em torno de $\langle k \rangle$ e a distribuição decai rapidamente. A Teoria dos Grafos torna-se uma ferramenta útil na compreensão de redes, principalmente aquelas de pequeno porte. A figura 2.7 ilustra um grafo aleatório.

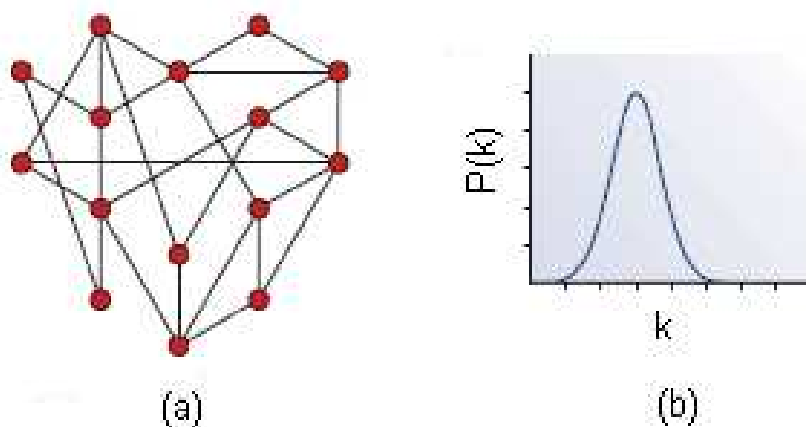


Figura 2.7: (a) Grafos aleatórios. (b) Distribuição do número de conectividades de cada sítio. Figura retirada da Ref. [6].

Uma propriedade dos sítios nas redes complexas é a tendência deles se agruparem formando “*clusters*”, ou seja, grupos de sítios fechados com conexões entre

vizinhos mais próximos. A agregação dos sítios na rede é caracterizada pelo coeficiente de agregação local dos sítios (C_i) [7] que é a razão entre o número das conexões entre os vizinhos mais próximos de um sítio i e o número total de ligações permitidas. Ou seja:

$$C_i = \frac{2n_i}{k_i(k_i - 1)} \quad (2.1)$$

k_i é o grau do sítio i e n_i é o número total de ligações entre os vizinhos mais próximos. Para um sítio i que possui k_i vizinhos, podem existir até $k_i(k_i - 1)/2$ ligações entre eles. O coeficiente de agregação, C , de uma rede é a média sobre todos os valores individuais C_i da rede. O valor do coeficiente de agregação fornece a probabilidade de encontrarmos uma conexão entre dois vizinhos de um sítio. Para modelos de grafos aleatórios, o coeficiente de agregação é $C = p = \frac{\langle K \rangle}{N}$, em que $\langle K \rangle$ é o número médio de conexões dos sítios e N é o tamanho da rede. Este coeficiente de agregação é válido desde que as conexões entre os sítios sejam distribuídas aleatoriamente. Ou seja, o coeficiente de agregação C é igual a probabilidade dos sítios estarem conectados. Em redes reais esse coeficiente de agregação C é muito maior que os valores encontrados para modelos com grafos aleatórios com os mesmos números de sítios e ligações.

Nas redes a distância entre dois sítios d_{ij} , representados por i e j respectivamente, é definida como o número de ligações ao longo do caminho mais curto que conecta esses dois sítios. A distância l é uma média sobre os pares de sítios da rede. Por exemplo, em uma rede de amigos l é o número de amigos em comuns que conectam duas pessoas da rede. Essa distância é calculada da seguinte forma,

$$l = \frac{1}{\frac{1}{2}n(n+1)} \sum_{i \geq j} d_{ij} \quad (2.2)$$

onde n é o número de sítios da rede e $\sum_{i \geq j} d_{ij}$ é a distância entre os sítios i e j . As redes reais possuem uma propriedade intrigante que é uma distância l pequena entre dois sítios, mesmo em redes grandes. Propriedade essa conhecida como “mundo pequeno”. Paul Erdős e Alfred Rényi [6] demonstraram que em grafos aleatórios a distância entre dois sítios é um função logarítmica do número de sítios na rede ($l \sim \ln(N)$).

Na teoria de redes há também o modelo de mundo-pequeno (small-world). Esse modelo foi proposto por Watts-Strogatz (Modelo WS) em 1998 [6]. Uma rede WS apresenta componentes de grafos regulares e grafos aleatórios. Um grafo é regular se todos sítios da rede possuem os mesmos números de conexões. Podemos considerar a rede WS como uma transição entre ordem e desordem. O modelo começa com um número n de sítios. Cada sítio é conectado aos seus K primeiros vizinhos em um grafo regular. Em seguida todas essas conexões são redirecionadas para outros sítios escolhidos ao acaso com probabilidade p , excluindo-se as autoconexões e as conexões duplas. No modelo WS temos a transição entre redes ordenadas ($p = 0$) e redes desordenadas ($p = 1$)(Figura 2.8). O modelo WS inicia-se com grafos regulares, mas em determinado momento emerge uma transição para os grafos aleatórios. A topologia apresentada no modelo WS descreve redes que podem ter tamanhos consideráveis, mas uma trajetória relativamente curta entre dois sítios quaisquer. O coeficiente de agregação para redes de mundo pequeno é dado por $C = \frac{3(K-2)}{4(K-1)}$ que converge para

$3/4$ no limite de K grande.

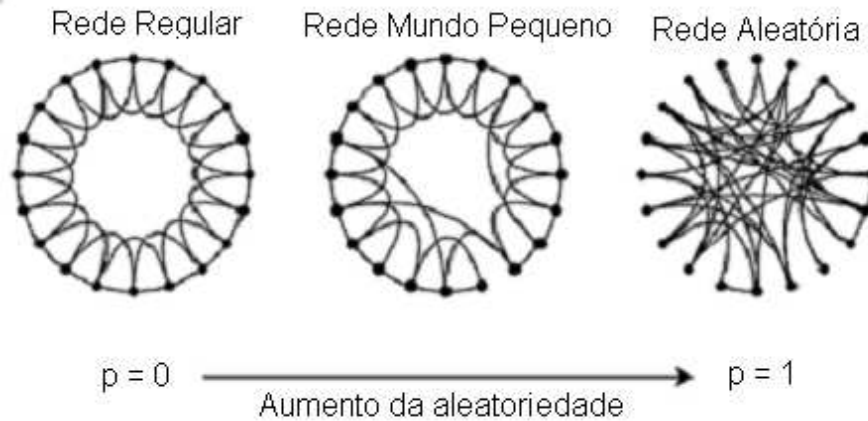


Figura 2.8: Modelo WS que inclui desde as redes regulares ($p = 0$) até redes aleatórias ($p = 1$). Figura adaptada da Ref. [6].

As redes complexas podem ser definidas como sendo estruturas que não apresentam padrões regulares. Em uma rede, primeiramente devemos determinar qual é o número de ligações de um sítio, grandeza que é denominada conectividade do sítio. Para um grafo regular, temos que cada sítio possui a mesma conectividade. Em um grafo irregular os sítios possuem conectividades diferentes. As redes complexas apresentam sítios com diferentes valores de conectividades. Portanto, os grafos regulares são casos particulares das redes complexas.

A topologia apresentada pelas redes complexas é um fator central nas análises dessas redes. Se tomarmos uma rede complexa que apresenta um tamanho pequeno, apenas a partir da análise visual, podemos inferir os aspectos centrais de sua estrutura e o peso relativo associado aos sítios e arestas dessa rede (Figura 2.9).

Entretando, se considerarmos uma rede com muitos sítios e com elevada conectividade, a análise visual, pouco ou quase nenhuma informação trará sobre a rede. A análise estatística assume agora um papel primordial no estudo da rede. Se observar-

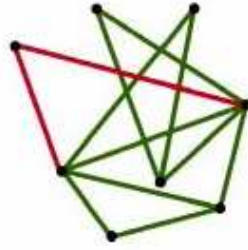


Figura 2.9: Rede livre de escala com um pequeno número de sítios e arestas. Figura adaptada da Ref. [7].

mos a figura 2.10, que traz a representação de uma rede de interação entre proteínas, uma análise estatística revela que a distribuição de conectividades dessa rede decai segundo uma lei de potência. A distribuição de probabilidades $P(k)$, fornece a probabilidade que um sítio, selecionado aleatoriamente, tenha k conexões. Essa distribuição de probabilidades $P(k)$ é dada da seguinte forma,

$$P(k) \sim k^{-\gamma}, \quad (2.3)$$

onde k é a conectividade de cada sítio e $k \in \mathbf{R}^+$. Ou seja, um reduzido número de sítios possui um elevado número de ligações. Esses sítios funcionam como hubs (sítios altamente conectados) da rede. Segundo Barabási [6], quanto mais conexões um sítio possui, maiores são as chances dele estabelecer novas conexões. Uma característica chamada de conexão preferencial. Um novo sítio tende a conectar-se com um sítio pré-existente que possui um maior número de conexões. As redes que seguem esse tipo de distribuição são denominadas redes livres de escalas. A evolução das redes no tempo com a inclusão ou remoção de sítios e arestas, as tornam ainda mais complexas e ricas. Toda essa complexidade pode emergir de regras de construção relativamente simples, como o mecanismo de conexão preferencial proposto por Barabási. Essa

topologia livre de escala produz valores menores da distância l entre dois sítios do que os valores correspondentes para redes aleatórias. Por sua vez, o coeficiente de agregação C apresenta valores superiores aos determinados para as mesmas redes aleatórias. Assim, a topologia livre de escala se apresenta como uma forma eficiente para organização das redes reais em razão da pequena distância entre os sítios e dos valores altos de agregação.

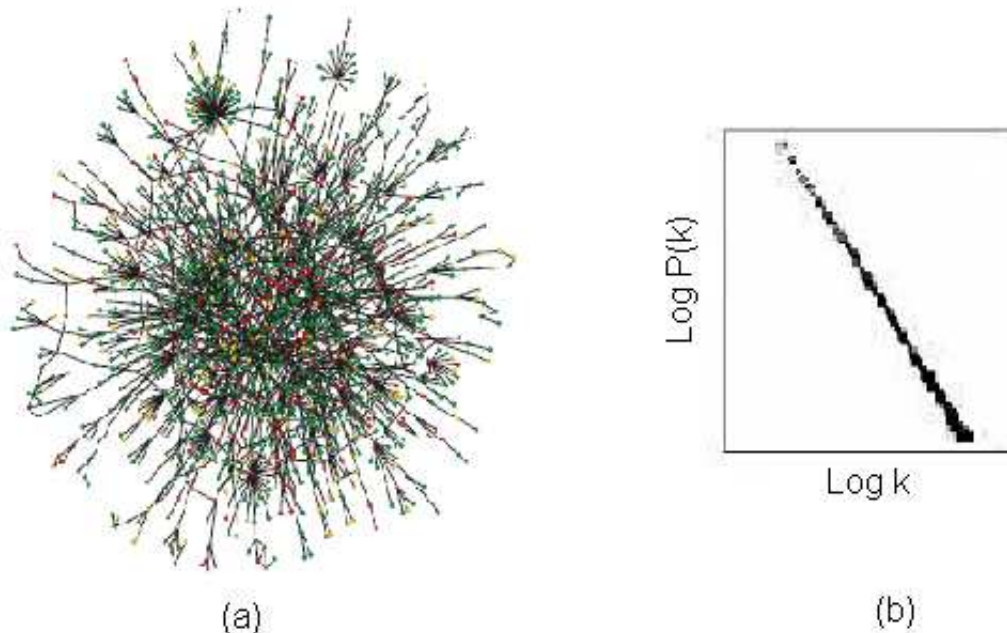


Figura 2.10: (a) Rede de interação de proteínas da levedura *Saccharomyces cerevisiae*. Os sítios são proteínas e as arestas são interações entre as proteínas. Há sítios com grandes números de conexões e outros com poucas conexões. (b) Distribuição de conectividade dos sítios. Figura adaptada da Ref. [8].

Um conceito importante em teoria de redes complexas é a assortatividade ou a desassortatividade que trata das conexões entre os hubs em uma rede. Por assortatividade entende-se o fato dos hubs em uma rede terem alta chance de se conectarem. Logo o conceito de desassortatividade refere-se a redes onde os sítios mais conectados, os hubs, não se conectam. Uma outra importante característica das redes complexas

com topologia livre de escala é a sua estabilidade. Elas são redes robustas que preservam as suas características mesmo diante de mudanças ambientais. Por exemplo, se ocorrerem remoções aleatórias de sítios, a probabilidade de remoção de um hub é baixa. Por outro lado, os sítios com baixas conexões têm maior probabilidade de serem removidos o que acarreta poucas mudanças na rede. Mas, se um ataque direcionado for feito aos hubs, a rede tende a se desmantelar. Essa robustez contra falhas é extremamente importante para o funcionamento de redes reais como, por exemplo, as redes biológicas e de comunicação. Por essa razão, as redes com estrutura livre de escala são tão comuns na natureza. As redes complexas podem também ser ditas direcionadas ou não. Ou seja, redes com segmentos específicos que possuem interesses afins, o que pode ocorrer em redes biológicas.

Em nosso trabalho utilizamos primeiramente o modelo livre de escala de Barabási e Albert (BA) que gera uma rede complexa caracterizada por uma distribuição de graus dos sítios que segue uma lei de potência com o expoente $\gamma = 3$. Redes com essas características têm sido observadas recentemente em redes metabólicas, internet, redes de contatos sexuais humanos, ou seja, redes reais [5]. A topologia das redes livres de escala no modelo BA é construída da seguinte forma:

(1) Inicia-se com um número pequeno de sítios (m_0) completamente conectados entre si, e a cada passo de tempo um novo sítio com m conexões ($\leq m_0$) é acrescentado à rede, ligando-se aos sítios já existentes; (2) A probabilidade Π de um novo sítio ser conectado a um sítio i já introduzido na rede, depende do número de conexão (k_i) de i e do número total de conexões existentes na rede. A probabilidade Π é dada pela distribuição:

$$\Pi(k_i) = \frac{k_i}{\sum_i k_i} \quad (2.4)$$

Portanto, quanto maior a conectividade de um sítio, maior será a probabilidade desse sítio receber uma nova conexão. Temos aqui um processo dinâmico de ligações preferenciais. Após $t - m_0$ passos de tempos, o algoritmo produz uma rede com N sítios.

As redes complexas descritas pelo mecanismo de BA estão recebendo uma grande atenção na literatura. Em um trabalho realizado por Barabási e Oltvar [8], a rede de interação entre proteínas da levedura *Saccharomyces cerevisiae* foi estudada. Nessa rede as conexões dos sítios apresentam distribuições com invariância de escala. A probabilidade $\Pi(k_i)$ de um sítio estar conectado a k outros sítios decresce segundo uma lei de potência caracterizada por um expoente λ . Para esse organismo foi encontrado $\lambda = 2.4$ [8].

O crescimento e as conexões preferenciais são os princípios gerais para poder-mos explicar as distribuições de conectividades em leis de potência ausentes nas redes de grafos aleatórios.

O segundo modelo de rede que adotamos, proposto por Aldana [9], também apresenta topologia livre de escala. A rede é construída da seguinte forma: N sítios fixos, e cada sítio i é conectado com os outros elementos da rede seguindo a distribuição de probabilidades $P_i(k)$ dada por,

$$P_i(k) = [\zeta(\gamma)k^\gamma]^{-1} \quad (2.5)$$

onde $\gamma > 1$ e $\zeta(\gamma) = \sum_{k=1}^{\infty} k^{-\gamma}$ é a função Zeta de Riemann.

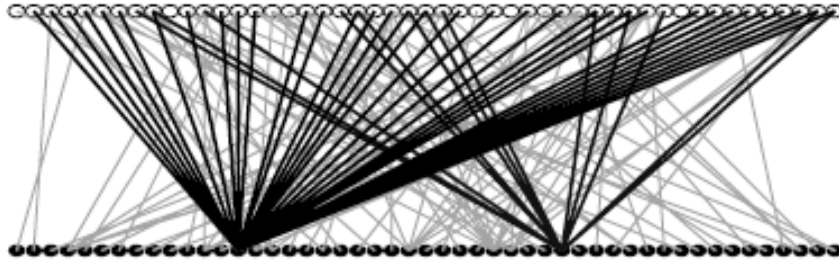


Figura 2.11: Esquema de uma rede com topologia livre de escala, para o caso de $\gamma = 2.5$. Os círculos pretos são os sítios da rede, enquanto os círculos brancos representam as entradas que atuam sobre esses sítios da rede. Figura retirada da Ref. [9].

Na figura 2.11 é mostrada uma representação de uma rede que obedece essa distribuição com $\gamma = 2.5$. O expoente livre de escala γ governa a transição de fase em nosso sistema. Exploraremos esse assunto no capítulo 4.

Apresentamos nas tabelas 2.1 e 2.2 algumas características de redes reais e como elas se comparam com seus valores correspondentes em redes aleatórias e livres de escalas do mesmo tamanho. A tabela 2.1 mostra o número N de sítios na rede, a conectividade média $\langle k \rangle$ dos sítios, a distância média de separação entre os sítios $\langle l \rangle$ e o coeficiente de agregação C da rede. Para comparação, nas duas últimas colunas estão incluídos $\langle l_{aleat.} \rangle$ e $C_{aleat.}$ para redes aleatórias com o mesmo tamanho e números de conexões. Na tabela 2.2, há algumas redes reais com topologia livre de escala com os seguintes parâmetros: tamanho da rede N , conectividade média $\langle k \rangle$, expoente da distribuição de conectividade γ e a distância média de separação entre os sítios $\langle l \rangle$. Na última coluna, temos $\langle l_{RLE} \rangle$ para a rede livre de escala.

<i>Rede</i>	N	$\langle k \rangle$	$\langle l \rangle$	C	$\langle l_{aleat.} \rangle$	$C_{aleat.}$
WWW	153127	35.21	3.1	0.1078	3.35	0.00023
Citações em artigos	70975	3.9	9.5	0.59	8.2	5.4×10^{-5}
Substratos de <i>E. coli</i>	282	7.35	2.9	0.32	3.04	0.026
Sinônimos de palavras	22311	13.48	4.5	0.7	3.84	0.0006

Tabela 2.1: Características gerais de algumas redes reais. Para cada rede são apresentados o número de sítios N , a conectividade média de cada sítio $\langle k \rangle$, o tamanho médio da menor distância entre dois sítios $\langle l \rangle$ e o coeficiente de agregação C . Para comparação com as redes reais, nas duas últimas colunas estão incluídos $\langle l_{aleat.} \rangle$ e $C_{aleat.}$ para redes aleatórias com o mesmo tamanho e números de conexões. Dados retirados da Ref. [6]

<i>Rede</i>	N	$\langle k \rangle$	γ	$\langle l \rangle$	$\langle l_{RLE} \rangle$
WWW	2×10^8	7.5	2.72	16	7.61
Atores de filmes	212250	28.78	2.3	4.54	4.01
Metabólica da <i>E. coli</i>	778	7.4	2.2	3.2	2.89

Tabela 2.2: Parâmetros de redes reais. Para cada rede temos, o tamanho da rede N , a conectividade média por sítio $\langle k \rangle$, o expoente da distribuição de conectividade γ e a distância média de separação entre os sítios $\langle l \rangle$. Na última coluna, temos $\langle l_{RLE} \rangle$ para a rede livre de escala para comparação com as redes reais. Dados retirados da Ref. [6]

2.4 As redes biológicas e o modelo de Barabási-Albert

Nesta seção discutiremos a adequação, pelo menos em primeira aproximação, do modelo de Barabási- Albert (BA) para descrever as redes biológicas. Também apontaremos as principais diferenças entre tais redes e o modelo livre de escala BA fundado em dois aspectos básicos: crescimento e conexão preferencial.

Uma vez que no interior de uma célula tudo encontra-se interconectado, especialmente as macromoléculas que interagem por meio de múltiplas ligações não covalentes, é natural tratá-la como uma rede. De fato, como um conjunto de pelo menos três redes emaranhadas que descrevem as interações proteína-proteína, proteínas-DNA e proteínas-metabólitos. Muitos resultados experimentais recentes, obtidos por meio

de técnicas de alto desempenho como cromatografia por afinidade, espectrometria de massa e análise de microarranjos, indicam que as redes de interações proteína-proteína, as rotas de catálise enzimática no metabolismo celular e os sítios de ligação para os fatores de transcrição que regulam a expressão gênica exibem características de redes livres de escalas [10–14]. Ainda mais, redes de transdução de sinais ou cascatas de sinalização e até redes neurais compartilham, a despeito de suas origens funcionais distintas, fortes similaridades em suas estruturas com as redes metabólicas, de interação proteína-proteína e de controle da expressão gênica.

Entretanto, algumas diferenças estatísticas importantes com relação às redes livres de escalas BA marcam estas redes biológicas de sinalização. Em especial, existem duas diferenças principais entre as duas. A primeira é a modularidade das redes que se reflete em elevados coeficientes de agregação C [15, 16] e gera uma correlação entre o coeficiente de agregação de um dado sítio e a sua conectividade k expressa por $c(k) \sim k^{-1}$ [17, 18]. Redes BA não exibem um espectro de agregação em lei de potência, o que indica a ausência de estruturas hierárquicas [19]. A segunda diferença relevante é a heterogeneidade da correlação grau-grau em redes biológicas, isto é, essa correlação obedece a uma lei de potência $\bar{k}_{nn}(k) \sim k^{-\nu}$ com $0 < \nu < 1$ [17]. \bar{k}_{nn} é o grau médio dos primeiros vizinhos de um nó com conectividade ou grau k . Esta relação entre os primeiros vizinhos do sítio não ocorre em rede livres de escala de BA [20].

O princípio de modularidade está disseminado nas redes biológicas. Sabe-se, por exemplo, que as proteínas funcionam em grupos co-regulados e levemente superpostos. Em analogia com a engenharia, um módulo em uma rede pode ser definido como um subconjunto de nós que interagem fortemente e possuem uma função

comum. Um módulo possui nós de entrada e de saída bem definidos que controlam a sua interação com o restante da rede. Assim, em um módulo existirão nós internos que interagem fracamente com outros nós externos. Mas, por que as arquiteturas modulares são abundantes em redes biológicas? Em engenharia, o uso de módulos é vantajoso nos casos em que as especificações de projeto mudam com o tempo. Os novos dispositivos podem ser facilmente construídos com base em módulos pré-existent e bem testados. Dispositivos não modulares nos quais cada componente conecta-se de forma ótima a outros componentes estão efetivamente congelados e são incapazes de evoluir para satisfazer novas exigências de otimização. Ao contrário, redes modulares podem ser prontamente reconfiguradas de forma a se adaptarem a novas condições. A centralidade deste argumento para a biologia é clara. A modularidade permite a adaptação de diferentes funções às custas de uma interferência reduzida em outras.

Diferenças à parte, o fato é que tanto o crescimento quanto a conexão preferencial, os dois princípios genéricos que explicam a emergência da topologia livre de escala, podem ser entendidas em redes biológicas. Por exemplo, redes metabólicas possuem uma história. Elas cresceram à medida que os organismos expandiram suas habilidades metabólicas e de biosíntese ao longo de bilhões de anos. O mesmo vale para as redes de proteínas cuja evolução, inferida a partir da comparação dos graus de similaridades entre os conjuntos de proteínas presentes em diversos organismos modelo, revela que as proteínas mais conectadas são as mais antigas e que o número de interações que uma proteína adquire ao longo dessa evolução é proporcional à sua conectividade [21]. Claramente, estes dois resultados estão em acordo com a dinâmica proposta por BA. De fato, o mecanismo de adesão preferencial, independentemente de

suas numerosas variações, sempre leva a este resultado central: nós altamente conectados são nós antigos, isto é, que foram adicionados muito cedo na história da rede. As proteínas antigas podem ser entendidas como os hubs em uma rede. Essas proteínas apresentam uma propriedade encontrada na teoria de redes que é a dessassortatividade, ou seja, as proteínas antigas não se conectam entre si. Propriedade esta que é consequência da modularidade presente nestes tipos de redes.

Porém ainda resta a pergunta: Por que uma proteína mais conectada é mais atraente para estabelecer novas interações? Dois mecanismos podem ser imaginados para responder em parte a esta questão. Primeiro, a evolução não apenas introduz novos nós, mas aperfeiçoa continuamente suas habilidades de interação. Quanto mais conexões um nó adquire, maior será a pressão seletiva para torná-lo mais conectado. Ao nível molecular haveria uma tendência para se aumentar o número de domínios de ligação desta proteína ou para melhorar os domínios já existentes e para torná-los capazes de se ligarem a mais alvos. A segunda possibilidade é que uma nova proteína ao modificar uma já existente, por exemplo, via fosforilação ou formação de complexos, pode melhorar ou regular aspectos da funcionalidade desta última. Portanto, a vantagem potencial de se permitir a modificação de uma proteína multifuncional seria aproximadamente proporcional ao número de funções distintas na qual ela está envolvida. Quanto mais a proteína está conectada, mais forte será a seleção em favor de se criar uma proteína que interage com ela.

A rede genética, que se encontra no núcleo celular, coordena todos os processos que ocorrem dentro da célula. A materialização deste fato está expressa no dogma central da biologia ($\text{DNA} \rightarrow \text{RNA} \rightarrow \text{Proteína}$) que enfatiza que as proteínas são codifi-

casas unicamente quando os genes são ativados. Essa rede ainda não foi integralmente desvendada ou modelada matematicamente. A grande diversidade dos constituintes das redes genéticas (os genes) e a complexidade das interações entre eles tornam difíceis a compreensão dos princípios organizacionais desses sistemas. Os modelos de redes livres de escalas, como os citados acima, parecem ser boas aproximações para o entendimento do funcionamento do genoma, pois estudos empíricos recentes [7, 22] têm demonstrado que a estrutura das redes genéticas dos eucariotos é do tipo livre de escala. Tal fato motivou o nosso trabalho com essa espécie de rede. A rede genética é organizada em primeira aproximação como descrevemos anteriormente ou seja, as ligações entre os genes (sítios) feita de modo probabilístico e de forma que um gene com maior número de interações possua uma probabilidade mais alta de interagir com novos genes. Essas ligações representam as interações bioquímicas entre os genes. O aparecimento de uma lei de potência para a distribuição de ligações entre os genes está intimamente atrelada à dinâmica de conexão preferencial entre os genes. Essa é uma propriedade que caracteriza a evolução de sistemas biológicos [10].

Capítulo 3

Modelos de Controle de Genes e de Diferenciação Celular.

Neste capítulo introduzimos alguns aspectos básicos dos autômatos celulares, do modelo de Kauffman, bem como o modelo de Sales-Martins-Stariolo, utilizado neste trabalho, e os principais resultados encontrados nesses modelos.

3.1 Autômatos Celulares

Os Autômatos Celulares (ACs) são modelos matemáticos construídos para descrever sistemas complexos nos quais o espaço e o tempo são discretos. Eles são compostos por um grande número de componentes simples e idênticos, chamados sítios ou células, com conectividade local (grafo com N sítios). Os ACs consistem geralmente de redes regulares d -dimensionais nas quais os sítios são variáveis discretas assumindo apenas um valor, dentre um conjunto finito de k valores possíveis. O estado do AC é especificado pelos valores de cada sítio. O novo valor associado a cada sítio depende do valor anterior desse sítio, bem como dos valores dos sítios conectados a ele, ou seja,

da sua vizinhança. O sistema evolui simultaneamente em passos de tempos discretos e de acordo com uma regra particular associada a cada sítio ou célula.

Uma das grandes vantagens dos ACs está em poder fazer de forma exata a sua manipulação em computadores digitais. Os ACs foram primeiramente introduzidos por von Neumann e Ulam como uma possível idealização de sistemas biológicos que se auto-reproduziam [23].

Partindo de estados iniciais aleatórios, sem nenhuma estrutura, um AC pode apresentar auto-organização com o surgimento de padrões espaço-temporais para tempos grandes. A regra de evolução do AC determina o padrão. A estrutura global emerge a partir de regras locais, uma característica de sistemas complexos [24].

A dinâmica dos ACs estudada por Wolfram na década de 80 tornou-se uma fonte de referência para diversos trabalhos na área dos ACs [23]. As regras unidimensionais de Wolfram ($d = 1$) determinam a evolução de um grupo de ACs que consiste de uma cadeia linear com N sítios. A cada sítio de uma rede com condições de contorno periódicas está associada uma variável $\sigma_i(t)$, o estado do sítio i ($i = 1, 2, \dots, N$) no tempo t . O valor do sítio pode ser qualquer inteiro pertencente ao intervalo $[0, k - 1]$: $\sigma_i = 0, 1, 2, \dots, k - 1$. O valor do sítio i mudará no tempo de acordo com uma regra determinística,

$$\sigma_i(t) = F_i[\sigma_{i-r}(t-1), \dots, \sigma_i(t-1), \dots, \sigma_{i+r}(t-1)] \quad (3.1)$$

em que F_i é uma função arbitrária que determina a regra do AC e r é o raio da vizinhança local do sítio i usada na regra. Dados os valores de r e k , temos k^{2r+1} possíveis

configurações para a vizinhança do sítio i . O espaço amostral das configurações dos ACs é dado por k^N e o número total de regras possíveis é $k^{k^{2r+1}}$. Apesar desse número enorme de regras de AC possíveis, Wolfram classificou os ACs em quatro classes de acordo com os padrões espaço-temporais gerados por sua evolução. Temos as seguintes classes de universalidade:

- i) classe I: A evolução leva o AC, após um período transiente, para um estado homogêneo (pontos fixos). (Figura 3.1(A));
- ii) classe II: A evolução leva o AC, após um período transiente, a um ciclo estável e periódico de configurações. (Figura 3.1(B));
- iii) classe III: A evolução leva o AC, após um período transiente, a formação de padrões caóticos. (Figura 3.1(C));
- iv) classe IV: A evolução leva à formação de estruturas complexas que possuem evolução imprevisível, uma situação intermediária entre o caos e a ordem (pontos fixos ou ciclos periódicos), sem análogo nos sistemas dinâmicos clássicos. (Figura 3.1(D)).

Na figura 3.1 mostramos os padrões gerados após a evolução de ACs com $k = 3$ e $r = 2$. As configurações iniciais eram desordenadas com cada sítio assumindo, independentemente, o valor 0, 1 ou 2 com probabilidade $1/3$.

3.2 O modelo de Kauffman

O modelo de Kauffman ou redes Booleanas aleatórias descreve sistemas dinâmicos desordenados [25]. Ele foi introduzido em 1969 para descrever os sistemas de

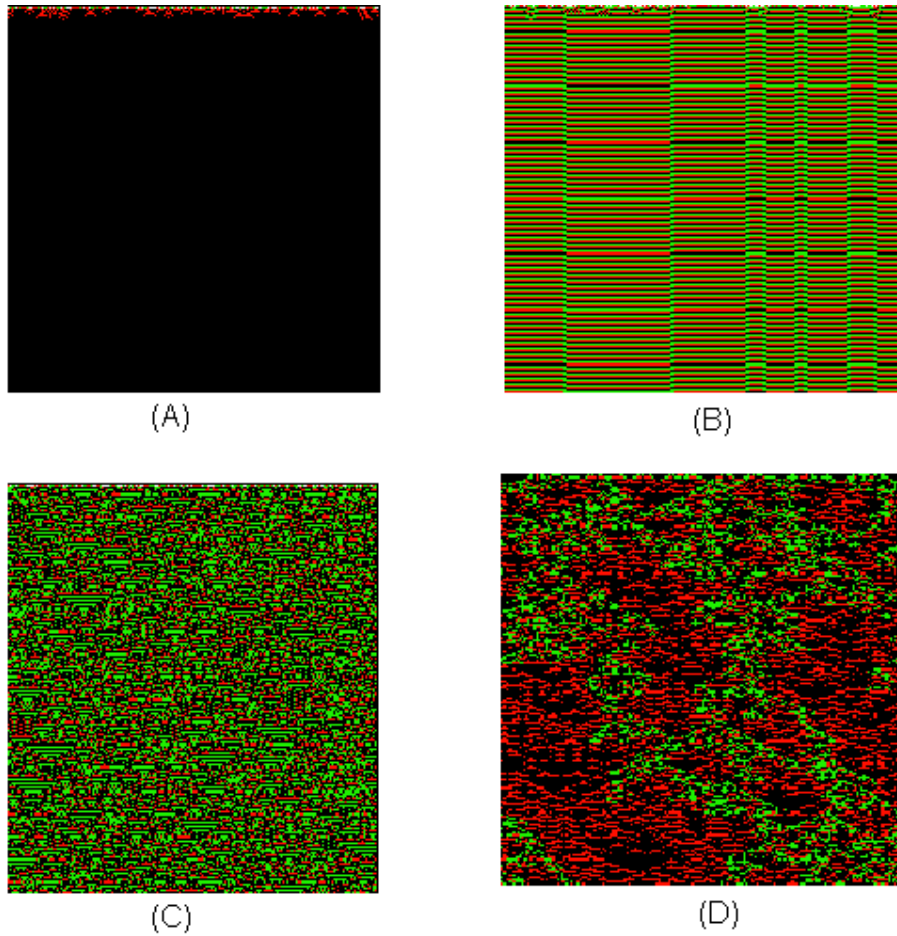


Figura 3.1: Padrões espaço-temporais das classes de Wolfram. (A) Classe I. (B) Classe II. (C) Classe III. (D) Classe IV.

regulação gênica que atuam na diferenciação celular. O modelo de Kauffman é definido da seguinte forma: o sistema consiste de N variáveis binárias, $\sigma_i(t) = \pm 1$ (+1 se o gene está ativo e -1 se ele está inativo para transcrição), $i = 1, \dots, N$. A atividade de um dado gene i é influenciada por K outros genes, $j_1(i), j_2(i), \dots, j_K(i)$, escolhidos aleatoriamente dentre os N genes. A evolução desse modelo em passos de tempo discretos é determinado por N funções Booleanas F_i , $i = 1, \dots, N$, de K entradas,

$$\sigma_i(t + 1) = F_i[\sigma_{j_1(i)}(t), \dots, \sigma_{j_K(i)}(t)] \quad (3.2)$$

escolhidas ao acaso dentre as 2^{2^K} funções Booleanas possíveis de K argumentos.

O AC de Kauffman, determinístico, não homogêneo e não local, é um exemplo de uma classe de modelos conhecidos como redes de autômatos aleatórios. A estrutura do grafo (rede) associado ao modelo de Kauffman é um grafo aleatório com N sítios e NK arestas orientadas que indicam as direções nas quais a informação flui. A aleatoriedade está na escolha dos K vizinhos de cada sítio e das funções F_i usadas para atualizá-los. Em resumo, a proposta original de Kauffman assume três pontos: (a) cada gene é controlado por K outros genes; (b) esses K genes são escolhidos aleatoriamente com probabilidade uniforme; (c) cada gene é inicialmente ativo para a transcrição de uma proteína com probabilidade p ou inativo com probabilidade $1 - p$.

Para fazer a interpretação biológica desse modelo, a variável σ_i representa o estado de ativação do gene i na célula. O gene examina os sinais que chegam das ligações (interações bioquímicas) com os outros genes da rede. O seu estado é ativado ou desativado de acordo com as regras de reação aos sinais enviados pelos K genes que o controlam. A rede evolui no tempo a partir de um estado inicial pela iteração repetida dessas regras. Ao longo de sua evolução, a rede genética pode passar por todos os seus possíveis estados, antes de repetir algum deles. Mas, normalmente, a rede atinge uma série de estados que ficam repetindo-se. Temos aqui um atrator para o qual um subconjunto de estados ou configurações possíveis convergem após um certo transiente. Esse subconjunto constitui a bacia de atração daquele atrator. As bacias de atração que se formam na dinâmica dessas redes Booleanas são interpretadas como os diferentes tipos celulares. Por sua vez, os períodos desses atratores são interpretados biologicamente como os comprimentos dos ciclos celulares, pois eles correspondem aos

tempos de recorrência dos padrões de atividade gênica na rede.

As propriedades dinâmicas dessas redes Booleanas aleatórias foram investigadas por vários autores [26] e, em resumo, os principais resultados encontrados foram: para redes com conectividade $K > 2$ os atratores são caóticos com baixa estabilidade à perturbações mínimas, os tamanhos dos ciclos crescem exponencialmente com N (aspectos indesejáveis na modelagem biológica) e o número médio de ciclos distintos é proporcional a N . Ao contrário, redes com $K = 2$ (a borda do caos) exibem uma ordem espontânea notável. Os tamanhos dos ciclos crescem com $N^{\frac{1}{2}}$ e tais ciclos são estáveis a quase todas as perturbações mínimas [25].

As redes Booleanas aleatórias de Kauffman podem espontaneamente exibir um comportamento dinâmico ordenado nas bordas do caos. No modelo original, a ordem e a estabilidade surgem quando emerge uma ilha percolante infinita de sítios congelados, ou seja, que permanecem com seus estados fixos (ativos ou inativos) ao longo da evolução temporal da rede. Por sua vez, os genes cujos estados oscilam entre ativação e inibição formam ilhas pequenas e isoladas. Toda essa estrutura de ilhas ocorre justamente quando a conectividade K do sistema atinge o valor crítico $K_c = 2$.

3.3 O modelo de Sales-Martins-Stariolo (SMS)

Esse é um outro modelo de autômatos celulares (AC) para estudar o comportamento de redes de genes que também utiliza variáveis binárias e regras de evolução simples. Como descrito na Ref. [27] o modelo representa um genoma por um conjunto de N genes binários distribuídos em uma rede, caracterizados por $\sigma_i(t)$, $i = 1, 2, \dots, N$. Para $\sigma_i(t) = 1$ o gene está ativo para a transcrição da proteína que ele codifica e

para $\sigma_i(t) = 0$ o gene está inativo e sua proteína não é sintetizada. O estado da rede em um tempo t é especificado pelo estado de cada sítio $\sigma_1(t), \sigma_2(t), \dots, \sigma_N(t)$. Cada gene i tem K sinais intracelulares (entradas) de regulação gênica. A auto-regulação ou auto-interação também faz parte desses sinais. As outras $K - 1$ entradas de cada gene i são escolhidos na vizinhança de Moore da rede quadrada subjacente com probabilidade p_2 e, com probabilidade $(1 - p_2)$, dentre qualquer outro sítio da rede fora dessa vizinhança local. Portanto, em sua versão original o modelo SMS utilizava uma rede do tipo SW ou small-world. A atividade dos genes em um passo de tempo é atualizada segundo a função,

$$\sigma_i(t + 1) = \Theta\left[J_{ii}\sigma_i(t) + \sum_{j=1}^{K-1} J_{j(i)}\sigma_{j(i)}(t) - \theta\right] \quad (3.3)$$

Portanto, o estado de um sítio i depende apenas dos seus elementos de regulação. O termo J_{ii} é a constante de acoplamento que representa a auto-regulação do gene i e $J_{j(i)}$ ($j = 1, 2, \dots, K - 1$) é a constante de acoplamento que representa a ação do sinal de entrada j sobre o gene i . A soma $(\sum_{j=1}^{K-1})$ estende-se sobre todos os sítios conectados com o gene i e θ é o limiar de ativação do gene i . No caso mais desordenado temos que $\theta = 0$. Por sua vez, $\Theta(x) = 0$ se $x \leq 0$ e $\Theta(x) = 1$ se $x > 0$, ou seja, a função degrau de Heaviside. Todos os genes são atualizados simultaneamente. Então, de acordo com a equação 3.3, um dado gene avalia o estímulo atual a ele conferido por todas as suas proteínas regulatórias, inclusive ele mesmo, e se este estímulo é positivo o gene se ativará ou permanecerá ativo (se ele já estava ativo) no próximo passo de tempo. Caso contrário, ele se tornará inativo ou permanecerá inativo para a transcrição.

Os acoplamentos $J_{j(i)}$ modelam o conjunto de interações bioquímicas complexas e parcialmente desconhecidas, característicos de redes metabólicas e genéticas. Esses conjuntos de interações bioquímicas complexas são responsáveis pelas taxas de síntese protéica. A escolha das constantes $J_{j(i)}$ foi feita de acordo com os seguintes aspectos biológicos. (i) os produtos de um determinado gene podem ativar, inibir ou simplesmente não afetar a transcrição de um outro gene. Nesse modelo todas as interações ativadoras assumirão um mesmo valor positivo $+J$ e as interações inibidoras $-J$. Quando o gene j não influencia a expressão do gene i , a constante de acoplamento será $J_{j(i)} = 0$. (ii) As interações entre os genes são assimétricas, isto é, $J_{ji} \neq J_{ij}$. O caso no qual um certo gene i ativa outro gene j , que por sua vez, inibe o gene i é biologicamente freqüente. (iii) A auto-regulação de um gene (controle autogênico) é também freqüente em organismos vivos. No modelo de Sales-Martins-Stariolo essa regulação é representada pelos acoplamentos J_{ii} .

Uma vez que os biólogos moleculares elucidaram apenas parcialmente a matriz de conectividade real entre os genes, os acoplamentos J_{ii} e J_{ij} são escolhidos de acordo com a distribuição de probabilidade.

$$P(J_{j(i)}) = \frac{(1 - p_1)}{2} [\delta(J_{j(i)} - 1) + \delta(J_{j(i)} + 1)] + p_1 \delta(J_{j(i)}) \quad (3.4)$$

onde $\delta(x)$ é a função delta de Dirac. Portanto, para uma rede de genes particular, cada constante J_{ji} pode ser ativadora (+1) ou inibidora (-1), com probabilidade $\frac{(1-p_1)}{2}$, e inexistente ou diluída (0) com probabilidade p_1 .

Ambos os modelos, o de Sales-Martins-Stariolo e o de Kauffman, são chamados

de modelos $N - K$, porque os N elementos que compõem o sistema interagem com outros K elementos que são escolhidos aleatoriamente.

3.4 Resultados básicos destes modelos

O modelo de Kauffman foi a primeira descrição teórica das redes de regulação gênica. Ele permitiu um entendimento introdutório da estrutura, da lógica e da dinâmica das redes genéticas. Como vimos, Kauffman introduziu uma rede Booleana aleatória, onde os genes eram do tipo on-off e atualizados ao mesmo tempo. Cada gene possuía K entradas regulatórias.

Os estudos teóricos do modelo de Kauffman mostraram que é conveniente generalizá-lo via a introdução de um parâmetro que regula a escolha das funções Booleanas. A estratégia mais utilizada na literatura, proposta originalmente em [28], é aquela de associar-se independentemente aos 2^K valores possíveis do argumento de F_i o estado 1, com probabilidade p , ou 0 com probabilidade $(1 - p)$. Agora, o espaço de parâmetros do sistema é constituído do par (K, p) .

Os trabalhos feitos desde a introdução das redes Booleanas têm demonstrado que elas apresentam dois regimes de comportamentos: congelado e caótico. Essas fases, nas quais o sistema pode se encontrar, descrevem como a informação flui no sistema. Em 1986 Derrida e Pomeau [28] mostraram analiticamente a existência dessas fases que são controladas pelos parâmetros K e p (probabilidade do gene estar ativo para a transcrição de uma proteína). Para cada valor de p há um valor crítico de conectividade $K_c(p) = [2p(1 - p)^{-1}]$. Para $K < K_c(p)$ o sistema encontra-se na fase ordenada e para $K > K_c(p)$ ele opera na fase caótica. De acordo com Kauffman

[9], quando $K = K_c(p)$ a fase crítica domina o sistema. A figura 3.2 apresenta as regiões ou fases do modelo de Kauffman em seu espaço de parâmetros. A região ordenada é preponderante para valores pequenos de K . Mas, em redes genéticas existe uma gama extensa de graus de conectividades. Por exemplo, a expressão do gene humano responsável pela produção de hemácias é controlado por mais de 20 proteínas regulatórias. Conforme pode ser visto na figura 3.2, para observarmos esse valor de K_c , o valor de p oscila entre uma alta probabilidade de ativação dos genes e um valor baixo de p . Esta é, portanto uma limitação no modelo de redes genéticas proposto por Kauffman.

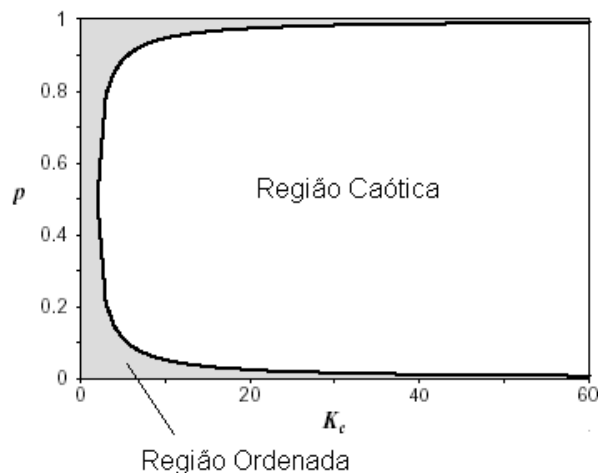


Figura 3.2: Diagrama de fases do Modelo de Kauffman com a região ordenada e a região caótica. Figura modificada da Ref. [9].

Na generalização da proposta original de Kauffman, a ordem caracterizada por uma ilha percolante infinita de genes congelados emerge em redes com mais de $K = 2$ entradas por sítio quando a maior parte das funções F são canalizadoras. O parâmetro p mede a tendência da saída da função Booleana ser 1, qualquer que seja a combinação de valores das suas K entradas. O parâmetro p é, assim, a fração das combinações de entradas que levam ao valor 1: o grau de canalização da regra Booleana. Acima de

um valor crítico $p_c(K)$ a rede Booleana aleatória estará no regime ordenado (Veja a figura 3.2). Para a rede quadrada, o valor crítico é $p_c = 0.72$

Um outro aspecto importante do modelo de Kauffman é a relação entre o número de atratores estáveis (os tipos celulares) no espaço de configurações possíveis da rede genética e o número N de genes. O número de atratores apresenta uma lei de potência em função do número N de sítios da rede na fase ordenada, o que concorda qualitativamente com a relação observada entre o número de tipos celulares de um organismo vivo e o seu número de genes. Na fase ordenada, essa relação cresce com $N^{2/3}$ [29]. Para pequenas redes há resultados numéricos que comprovam essa lei de escala [30]. Na fase caótica o número de atratores em função do número N de genes apresenta um crescimento exponencial. Kauffman interpreta cada atrator como um tipo celular. Cada estado do sistema no seu modelo representa um padrão instantâneo de atividade gênica da célula. Assim, a célula começa com um padrão inicial de expressão que evolui para um tipo celular particular, o atrator. Biologicamente, nenhum organismo vivo pode estar baseado em ciclos de períodos muito longos. A seleção a partir de mutações exige um fenótipo recorrente sobre o qual agir e, portanto, um sistema genômico cujos padrões de atividade gênica se repetam em ciclos curtos. Na fase caótica o tamanho dos ciclos cresce exponencialmente com o tamanho N da rede genética. Assim, o sistema percorre muitos estados antes de alcançar uma configuração estável. Em contrapartida, na fase congelada os ciclos atratores são mais curtos que na fase caótica, obedecendo a uma lei de potência. Um conjunto inicial de genes, rapidamente chega em uma configuração estável. Na fase crítica do modelo de Kauffman há a estabilidade compatível com sistemas biológicos reais, ou seja, ordem, flexibilidade e estabilidade

frente a mutações que são necessárias a um organismo vivo. Nessa fase a rede genética que representa a célula pode alcançar rapidamente um atrator estável.

No modelo de Sales-Martins-Stariolo [27], foram usados $K = 9$ entradas regulatórias por genes, incluindo a auto-interação, e a estrutura de rede adotada foi a de mundo-pequeno (small-world). Os genes regulatórios em organismos procariotos e eucariotos são, em média, diretamente controlados por até 6 ou 10 proteínas. As $K - 1$ entradas são escolhidas aleatoriamente entre os outros genes restantes da rede com probabilidade p_2 (probabilidade de um acoplamento na vizinhança de Moore), ou são genes com acoplamento de longo-alcance com probabilidade $1 - p_2$. O limite $p_2 = 1$, corresponde a um modelo de alcance infinito com grau $K = 9$ (incluindo a auto-interação J_{ii}). Já o limite $p_2 = 0$ corresponde a uma rede quadrada na qual cada sítio tem uma vizinhança de Moore, definido pelos oitos vizinhos mais próximos. Para qualquer outro valor de p_2 existem simultaneamente na rede acoplamentos de alcances curto e longo, refletindo o fato biológico que um dado gene poder ser regulado ou por uma seqüência de DNA próxima ou por uma seqüência distante na molécula ou até mesmo em outro cromossomo. Como já mencionado, do ponto de vista da teoria de grafos, as redes do modelo SMS correspondem ao tipo WS.

Os resultados apresentados na referência [27] correspondem a redes genéticas com tamanho máximo de $N = 625$ genes e estados iniciais dos genes igualmente prováveis. A existência de uma fase marginal entre as fases congelada e caótica é o resultado central desse trabalho. O regime congelado apresenta a maioria dos genes fixos no estado ativo ou inativo e poucos genes oscilam entre esses dois estados. Esse é um sistema estável onde as perturbações introduzidas no sistema não afetam os

estados dos genes. O regime caótico apresenta os genes em sua maioria oscilando entre os estados ativos e inativos. O sistema é instável a pequenas perturbações.

A fase marginal marca a transição entre a ordem congelada, inflexível e o caos. Como já citamos anteriormente, a borda do caos. Nessa fase o genoma exibe sensibilidade moderada às condições iniciais. A percolação dos genes congelados nessa fase é suficiente para ordenar o comportamento da rede. Na figura 3.3 apresentamos o diagrama de fases desse modelo. No modelo de Kauffman a fase ordenada é excessivamente rígida em comparação com o regime marginal encontrado por Sales-Martins-Stariolo. Estes autores também interpretam os vários tipos celulares como ciclos estáveis que emergem da interação dos genes com seus vizinhos. O número limitado de ciclos atratores representa o repertório restrito de tipos celulares que podem ser gerados pela diferenciação celular em organismos vivos. Os períodos desses atratores podem ser associados à duração dos ciclos celulares percorridos pelas células em mitose que devem replicar todos os seus genes. Para o modelo de Sales-Martins-Stariolo na região caótica, o comprimento dos ciclos atratores apresenta um crescimento exponencial com N , enquanto que nas regiões congelada e marginal uma lei de potência em N foi observada. Além disso, na região marginal desse modelo os padrões de expressão gênica são estáveis a perturbações, permitindo assim que mutações ocorram e, portanto, que a evolução natural aconteça no sistema. Conseqüentemente, a evolução Darwiniana pode se manifestar por meio da acumulação gradual de pequenas mutações nessa região.

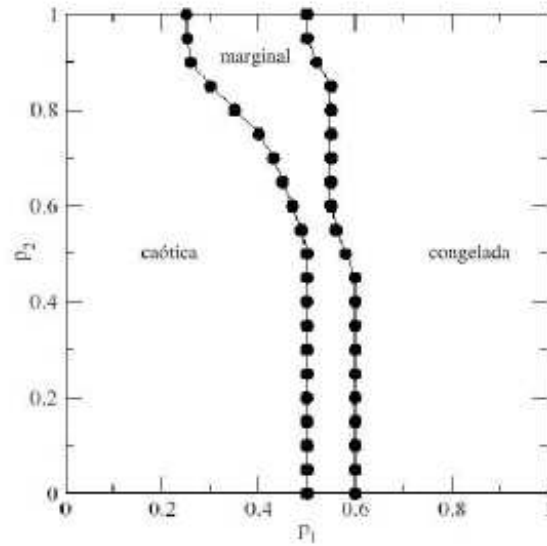


Figura 3.3: Diagrama de fases do Modelo de Sales-Martins-Stariolo com três regiões. Nesse diagrama foram usados $N = 400$ genes e 1000 amostras. Figura retirada da Ref. [27].

Capítulo 4

O Modelo de Sales-Martins-Stariolo em Redes Livres de Escalas

Nesse capítulo trataremos da estrutura dos atratores e da propagação de danos no modelo Sales-Martins-Stariolo. Vamos apresentar também os resultados que obtivemos e compará-los com aqueles relativos ao modelo de Kauffman.

4.1 Estrutura de atratores

A análise da dinâmica de um determinado AC parte da observação dos padrões espaço-temporais que ele gera. Sistemas dinâmicos como os ACs armazenam informações em seus atratores. Na dinâmica dos ACs determinísticos uma configuração $\sigma_i(t+1)$ depende unicamente da configuração prévia $\sigma_i(t)$. Para um AC finito com N sítios a evolução temporal sempre se tornará periódica, pois o seu espaço de configurações é finito ($\Omega = k^N$ estados diferentes possíveis). O número de estados possíveis do espaço de configurações cresce exponencialmente com N . Portanto, após

um tempo $t > k^N$ o AC deverá, necessariamente, ter visitado pela segunda vez uma mesma configuração e, a partir de então, percorrer repetidamente uma seqüência de estados. Por exemplo, para um AC com $k = 3$ estados e $N = 10$ sítios existirá um total de $k^N = 3^{10} = 59049$ configurações possíveis no espaço de estados. Os ciclos presentes nos ACs são os atratores periódicos da dinâmica do sistema. Todas as configurações que convergem para um dado atrator constituem a sua bacia de atração. Para N finito, o espaço de estados pode se dividir em várias bacias de atração correspondentes a atratores distintos. Na figura 4.1 há uma representação das configurações de bacias de atração. Cada condição inicial do sistema vai eventualmente para um desses ciclos, que são determinados pelas regras de evolução do sistema em estudo. O conceito de estados transientes na teoria dos ACs, relaciona-se com as regiões ou subconjuntos de configurações que estão fora do atrator.

Temos duas quantidades importantes que caracterizam os atratores no espaço de configurações: o número de atratores diferentes e o período ou comprimento do atrator. Conseqüentemente, o estudo do AC pode ser feito pela descrição quantitativa das seguintes grandezas: distribuições dos períodos dos ciclos dos atratores, dos tamanhos das bacias de atração e dos transientes.

4.2 A propagação de danos

Em nosso trabalho, no qual estudamos o autômato celular proposto por Sales-Martins-Stariolo em uma rede livre de escala, uma perturbação (dano) foi introduzida como forma de sondarmos o comportamento de nosso sistema. Esta análise de estabilidade dos padrões espaço-temporais gerados pela evolução do sistema com relação a

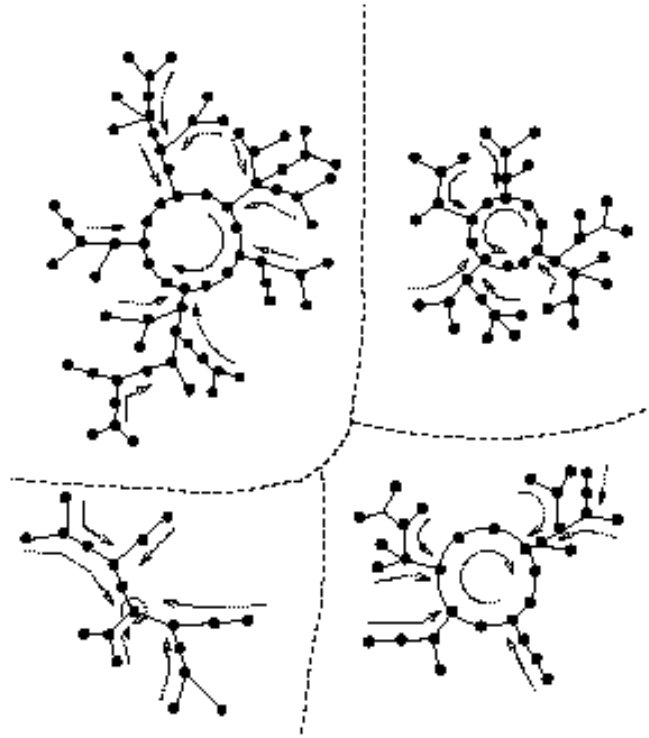


Figura 4.1: Esquema de partição do espaço de estados por seus atratores. Cada estado está sendo representado por um ponto. Os atratores são representados como ciclos. Os estados que evoluem para um ciclo, seguindo as setas, formam uma bacia de atração. Figura retirada da Ref. [9].

pequenas perturbações, introduzidas no estado inicial, é uma forma de se caracterizar as classes de ACs. Nela, utilizamos dois sistemas: o original e uma réplica na qual uma pequena perturbação é introduzida. A configuração inicial da réplica é criada a partir da configuração inicial do sistema original com alguns poucos sítios, escolhidos aleatoriamente, em estados alterados. Por exemplo, de ativos para inativos. Então, faz-se a evolução dos sistemas e compara-se a trajetória das configurações original e da réplica perturbada. O efeito global dessas perturbações pode ser medido através da distância de Hamming ($\Psi(t)$) que é definida como o número total de sítios cujos valores diferem nessas duas configurações após t passos de tempo. Matematicamente,

$$\Psi(t) = \frac{1}{N} \sum_{i=1}^N | \sigma_i(t) - \sigma'_i(t) | \quad (4.1)$$

onde N é o número de sítios do AC, $\{\sigma(t)\}$ e $\{\sigma'_i(t)\}$ são dois padrões associados ao sistema original e a réplica, respectivamente, que se diferenciam no tempo $t = t_0$ em uma fração $\Psi(t_0)$ de sítios. Normalmente essa fração $\Psi(t_0)$ de sítios perturbados é bem pequena. Se o sistema é fortemente sensível às condições iniciais, a perturbação introduzida propaga-se por toda a sua extensão. A região na qual os dois padrões são diferentes se expande sem limite no tempo. A extrema sensibilidade às condições iniciais é um dos aspectos básicos dos sistemas caóticos. Em particular, em um sistema caótico uma pequena perturbação pode levá-lo para uma bacia de atração totalmente diferente da original.

Analogamente à classificação citada no capítulo 2, as quatro classes de Wolfram podem ser distinguidas com base no comportamento da distância de Hamming:

- i) classe I: A distância de Hamming inicial desaparece, $\Psi(t \rightarrow \infty) = 0$, independentemente de $\Psi(0)$. O padrão espaço-temporal gerado é absolutamente estável;
- ii) classe II: A distância de Hamming permanece constante no tempo com um pequeno valor e localizada. Padrões de evolução estáveis;
- iii) classe III: A distância de Hamming aumenta com o tempo e atinge um valor de equilíbrio alto. Mesmo com um dano inicial pequeno, ela propaga-se por toda a rede. Essa classe é instável e exhibe o comportamento caótico;
- iv) classe IV: A distância de Hamming inicial não desaparece, porém permanece pequena. Há um período transiente caótico antes que o valor da distância de Hamming

seja atingido.

O AC está em uma fase ordenada se, para um dado valor de N e um tempo t suficientemente longo, a distância de Hamming $\Psi(t)$ tem um valor pequeno ou desaparece. A presença de uma distância de Hamming finita não-nula para $\Psi(0) \rightarrow 0$ que se propaga por toda a rede, demonstra que o AC está em uma fase caótica. Diz-se, nesse caso que a evolução da dinâmica do AC é extremamente sensível às condições iniciais. Para estimarmos com mais precisão o limite $\Psi(0) \rightarrow 0$, em nossas simulações numéricas utilizamos a estratégia, tal como proposto por Sales et al. [27], na qual são considerados três estados iniciais σ_1 , σ_2 e σ_3 , com $\Psi_{12}(0) = \Psi_{23}(0) = \frac{1}{2}\Psi_{13}(0)$. Aqui Ψ_{12} é a distância de Hamming entre as configurações σ_1 e σ_2 e Ψ_{23} , Ψ_{13} a distância entre as outras configurações. A distância de Hamming calculada da seguinte forma,

$$\Psi(t) = \Psi_{12}(t) + \Psi_{23}(t) - \Psi_{13}(t) \quad (4.2)$$

fornece uma extrapolação mais precisa para $\Psi(0) \rightarrow 0$.

4.3 Resultados e Discussão

Ao simularmos o modelo de Sales-Martins-Stariolo, utilizamos primeiramente a rede livre de escala proposta por Albert-Barabasi (discutida na seção 2.4). Na rede, variamos o número de conexões iniciais por gene (m) de 1 a 6 e cada gene tinha igual probabilidade (50%) de estar ativo ou inativo no estado inicial. O espaço de estados desse modelo possui uma gama enorme de estados iniciais. Utilizamos 10 estados iniciais nas simulações computacionais, sorteados aleatoriamente, 100 redes distintas e

o número de genes utilizado foi $N = 2048$ para uma máxima desordem ($\theta = 0$). Nessas simulações o dano cresce rapidamente com o tempo e atinge um valor de equilíbrio que depende do número de conexões m que cada gene faz ao entrar na rede. Por exemplo, com apenas um sítio perturbado na rede genética o dano propagava-se no sistema. Esse comportamento típico da distância de Hamming ($\Psi(t)$) é mostrado na figura 4.2.

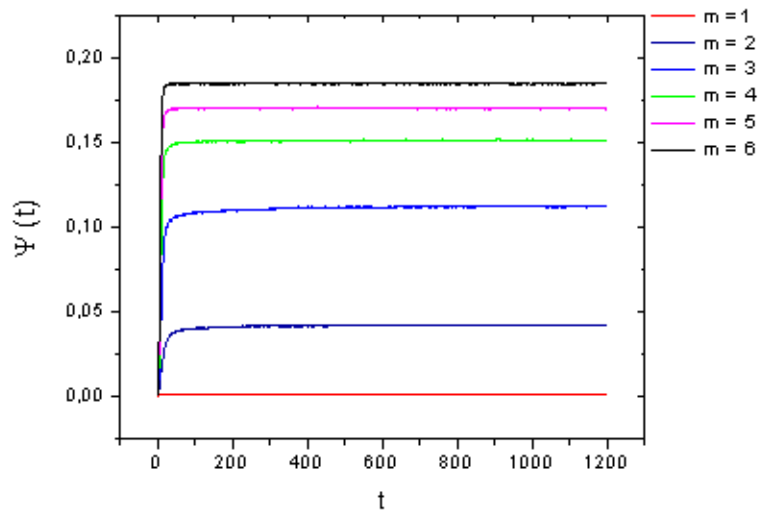


Figura 4.2: Distância de Hamming em função do tempo. Um dano inicial da ordem de 0.1% de genes perturbados foi introduzido na atividade genética. Os dados correspondem a 100 redes diferentes, com $N = 2048$ e $\theta = 0$ (Máxima desordem). Espera-se que, no limite $N \rightarrow \infty$ e $t \rightarrow \infty$, as flutuações no gráfico diminuam.

Para $m = 6$, o maior número de conexões iniciais por gene que adotamos, a distância de Hamming apresentou um valor de equilíbrio superior aos demais. Ou seja, para uma rede genética com esse número de conexões iniciais, apenas um gene perturbado pode gerar um dano em cerca de 18% da rede genética. Isso equivale a uma única mutação que pode transformar um organismo biológico, como por exemplo um rato em um homem. De fato, a diferença genética entre homens e os grandes

símios é de apenas 1%, aproximadamente. Portanto, o sistema encontra-se em uma região caótica. Em sistemas onde a conexão inicial por sítio é $m = 2$, a perturbação final é cerca de 4% da rede. Uma propagação de danos ainda alta para sistemas biológicos. Já com $m = 1$ o sistema apresenta-se em uma fase congelada onde o dano não se propaga.

O efeito do tamanho das perturbações iniciais introduzidas no sistema é analisado na Figura 4.3. Introduzimos o dano inicial ($\Psi(0)$) no passo de tempo $t = 0$ e fizemos uma média sobre $\Psi(t)$ após t passos de tempo a partir de $t = 1000$.

Para $m = 1$, os genes encontram-se praticamente congelados. A distância de Hamming final cresce muito lentamente com a distância de Hamming inicial. O dano introduzido na rede genética não tende a se propagar pelo sistema. Uma mutação que atinge um determinado gene não afeta uma grande quantidade de outros genes do sistema. Nesse regime congelado a diferenciação celular tem pouco, ou nenhuma probabilidade de ocorrer. Analisando a curva que representa o sistema para $m = 6$, a distância de Hamming final é alta, independente do dano inicial. As mutações proliferam pelo sistema. A percentagem de genes que sofreram alterações pode chegar a 20% da rede genética e o sistema está no regime caótico. Uma curva intermediária entre a alta rigidez de um regime congelado e a desordem do regime caótico pode ser observada para $m = 2$. Os três regimes, mencionados na seção 3.4, são determinados pela variação da conexão inicial (m) com a qual cada gene entra na rede. O regime marginal possui as características necessárias para a adaptação, a auto-organização de sistemas complexos e, portanto, para a evolução dos organismos vivos. As nossas redes livres de escalas com $m = 2$ apresentaram a conectividade média $K_{média} = 4.02$. Para

redes com outros valores de m a conectividade média é proporcional a m ($\langle k \rangle \sim 2m$). No modelo de Kauffman [25] o limite da fase ordenada ocorre para uma conectividade média entre genes $K_c = 2$. Já Sales et al. [27] encontraram que a região marginal se estende até $K_{média} = 5.4$.

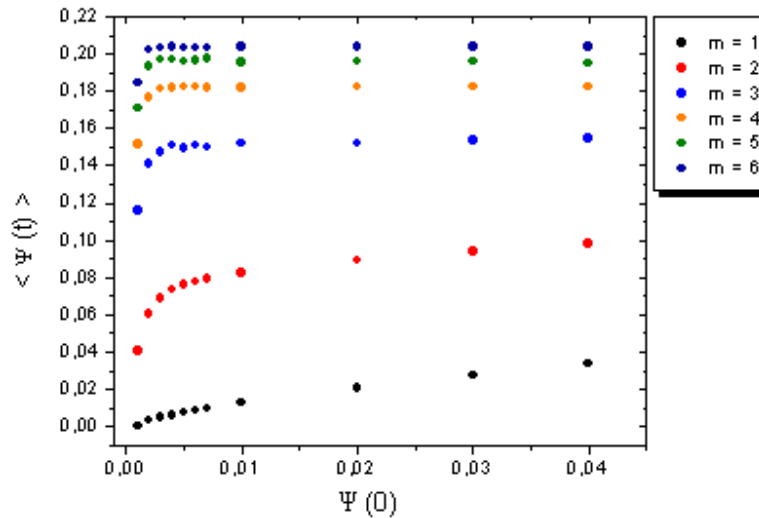


Figura 4.3: Dano final médio em função do dano inicial para diferentes valores m de ligações iniciais por gene e $\theta = 0$. Estes resultados correspondem a redes genéticas com $N = 2048$ genes e 2000 amostras diferentes.

O parâmetro θ no modelo de Sales-Martins-Stariolo é um limiar constante que força os genes para o estado de inibição. Fizemos diversas simulações computacionais variando esse parâmetro, com o intuito de observar a influência desse limiar sobre o comportamento do sistema. Para cada valor de θ adotado (variação de 0 a 2) construímos os gráficos de $\Psi(t) \times \Psi(0)$. O aumento do valor de θ tende a diminuir a distância de Hamming final. Na figura 4.3 temos $\theta = 0$, que quando comparado à figura 4.4 ($\theta = 0.8$ e $m = 6$), exibe uma distância de Hamming estacionária superior.

Para determinarmos o valor da distância de Hamming efetiva, para cada valor

de θ adotado, utilizamos,

$$\Psi_N = \lim_{\Psi(0) \rightarrow 0} \Psi_\infty(\Psi(0)) \quad (4.3)$$

Assim, à distância de Hamming final média é um limite para o caso no qual a distância de Hamming inicial tende a zero. Ou seja, à distância de Hamming para a menor perturbação possível introduzida no sistema.

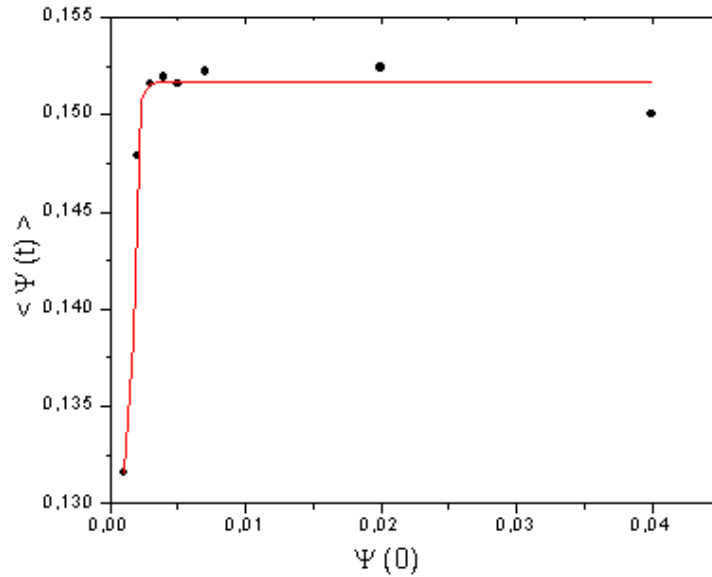


Figura 4.4: Dano final médio em função do dano inicial para $\theta = 0,8$, $m = 6$, $N = 2048$ genes e 2000 amostras diferentes.

A variação da distância de Hamming final média com θ é apresentada na figura 4.5. Como esperado, o dano Ψ_N tende a diminuir à medida que o valor de θ aumenta. De fato, o limiar de ativação (θ) ordena o sistema, de tal forma que o dano propaga-se com mais dificuldade à medida que θ aumenta. Logo, θ força os genes para um estado congelado. O dano médio final Ψ_N é, como se vê na figura 4.5, uma função seccionalmente constante do limiar de ativação θ . Ao variarmos o parâmetro de ativação θ ,

apenas em valores inteiros de θ , temos uma diminuição de Ψ_N . Variando-se θ , pode-se passar de um regime caótico, com elevado valor de Ψ , até um estado congelado. Assim, o limiar θ desempenha um papel similar à probabilidade p no modelo de Kauffman generalizado e é um parâmetro que controla as transições de fase no modelo de Sales-Martins-Stariolo. Quanto maior é o valor de θ , mais ordenado o sistema tende a permanecer.

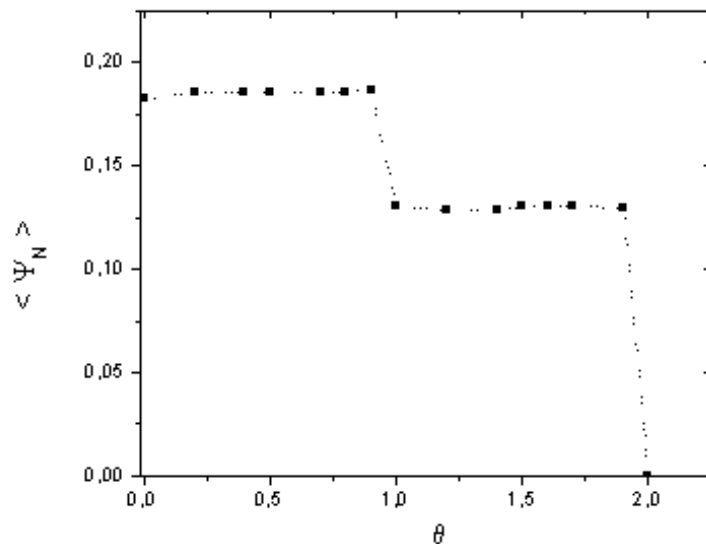


Figura 4.5: Dano final médio em função de θ . Os dados correspondem a redes genéticas com $N = 2048$ genes, número de conexões iniciais $m = 6$ e 200 amostras diferentes.

No espaço de parâmetros constituídos pelo limiar de ativação dos genes (θ) e pelo número de conexões iniciais de cada gene (m), analisamos o comportamento da propagação de danos no modelo. Como visto nas figuras 4.3 e 4.5, a conectividade inicial por gene (m) e o limiar de ativação θ afetam o modo como o dano (mutação) propaga-se através da rede genética, sendo assim m e θ são então os parâmetros relevantes na análise das regiões do diagrama de fase desse modelo. O diagrama

de fases para o modelo de Sales-Martins-Stariolo na rede livre de escala de Albert-Barabási exhibe uma região congelada e uma região caótica, tal como proposto por Kauffman. A Figura 4.6 apresenta essas fases de acordo com os valores dos parâmetros θ_c e m . Adotamos o valor de θ_c como sendo aquele a partir do qual a distância de Hamming Ψ_N vai à zero na rede de genes. Ou seja, o valor de θ_c força ou mantém os genes a operarem em um regime congelado.

No segundo modelo de redes livres de escalas que adotamos foi realizada a variação do expoente γ que controla a distribuição do grau de conectividade $P(k)$ da rede. Esse expoente é um novo parâmetro que afeta as transições de fase que ocorrem na rede. Na figura 4.7, temos um gráfico das distribuições de conectividades dos sítios de uma rede com $N = 80000$ genes e $\gamma = 2$, apresentando um comportamento livre de escala. As redes livres de escalas que descrevem comportamentos de sistemas reais, tais como sistemas biológicos, sociais, ecológicos e outros apresentam expoentes γ que estão numa faixa entre 2 e 2.5 [9].

Na figura 4.8, estão mostradas as distâncias de Hamming (Ψ_N) para alguns valores fixos do expoente γ em função do limiar de ativação θ . O dano inicial foi introduzido na rede no passo de tempo $t = 0$. Como se pode ver, quanto maior o valor do expoente γ , menor é a perturbação que se propaga pela rede genética.

Essa redução no valor do dano assintótico com o aumento de γ , deve-se a distribuição de conectividade associada à rede. Como pode ser observado na figura 4.9 (distribuição de conectividade para diferentes valores de γ), a curva mais inclinada (maior γ) tem uma menor probabilidade do dano ser gerado em genes mais conectados. Sendo assim o dano não tende a ser propagar para outros genes da rede. Ao contrário,

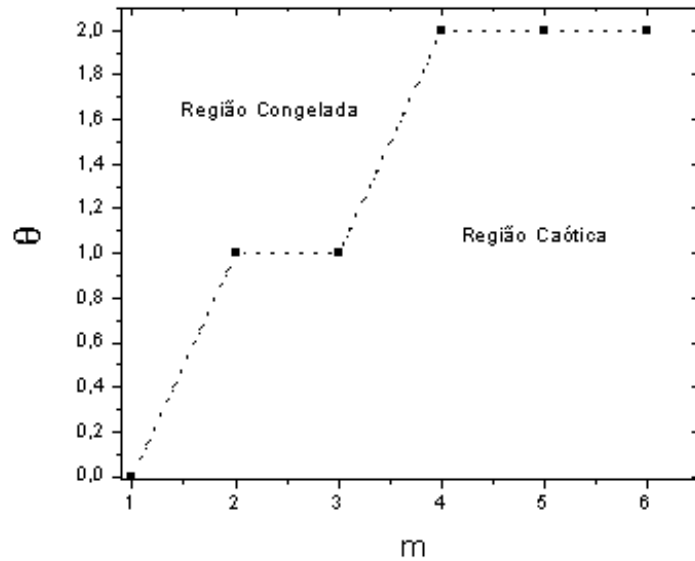


Figura 4.6: Diagrama de fases do modelo Sales-Martins-Stariolo na rede livre de escala de Albert-Barabási. As fases são delimitadas em função do limiar de ativação crítico (θ_c) e da conectividade inicial de cada gene m . Os dados correspondem a redes genéticas com $N = 2048$ genes e 200 amostras.

um valor baixo de γ gera uma rede genética com conectividades mais uniformes. A perturbação, a mutação genética, é distribuída de maneira mais igualitária no sistema. Logo, uma rede genética com $\gamma = 2.4$ e $1 \leq \theta < 2$ apresenta uma taxa de propagação de mutações que pode ser compatível com o que se espera nos organismos vivos. Por exemplo, em uma rede genética com 30.000 genes, o número de genes codificadores de proteínas encontrado nos núcleos das células do ser humano, é biologicamente possível que uma mutação ataque cerca de 100 outros genes [31], o que equivale a cerca de 0.33% de perturbação na rede. Para a nossa rede com $\gamma = 2.9$, temos cerca de 0.9% de genes alterados para o estado mais desordenado.

O diagrama de fases com os parâmetros θ_c e γ (Figura 4.10) exhibe uma região caótica e uma região ordenada. Assim, por exemplo, para $\gamma = 2$ e $\theta > \theta_c = 3$ ou para

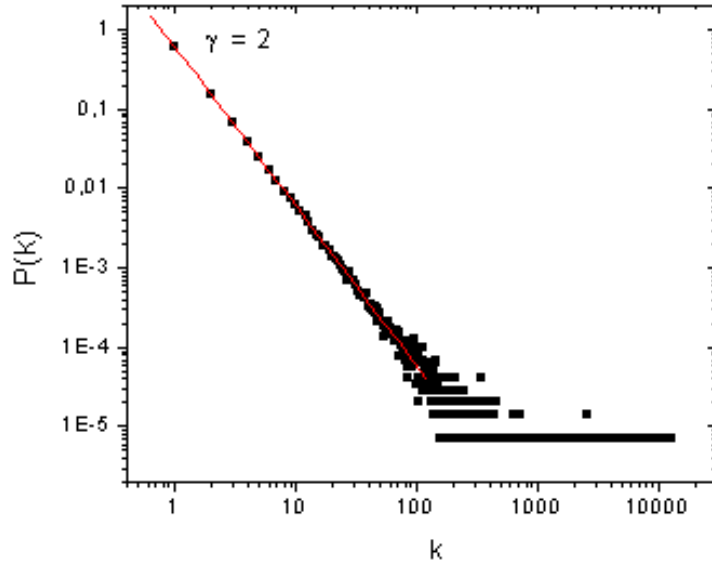


Figura 4.7: Gráfico da distribuição de conectividades dos sítios de uma rede genética com $N = 80000$ genes e $P(k) \sim k^{-\gamma}$ com $\gamma = 2$. Foram geradas 10 redes diferentes para a realização das médias.

$2.1 \leq \gamma \leq 3$ e $\theta > \theta_c = 2$ a rede de genes está na fase ordenada.

Na região congelada a maioria dos genes está no estado ativo ou inativo. Os períodos dos atratores são pequenos, todos com praticamente o mesmo valor (Figura 4.11), crescendo linearmente com N . As alterações (perturbações) aplicadas nos genes não se propagam pelo sistema, ou seja, temos uma dinâmica estável. Na região caótica os atratores apresentam períodos que crescem mais rapidamente de forma linear. Pequenos danos se propagam para uma fração finita dos genes do AC e, portanto, o sistema é sensível às condições iniciais. Devemos salientar que o tamanho pequeno da rede pode ser o responsável por esse comportamento dos períodos dos atratores, bem como as características dos atratores em função do tamanho N da rede que serão apresentadas logo abaixo.

Os tipos celulares diferentes (atratores) nessas redes livres de escalas apresen-

taram comportamentos distintos dos observados por Aldana [9]. No seu modelo, para a região caótica o número de atratores cresce linearmente com o número N de genes. Veja a Figura 4.12. Em nosso modelo na fase ou regime caótico, os atratores crescem exponencialmente com N (Veja a Figura 4.13). Como era de se esperar, nessa região caótica há extrema sensibilidade às condições iniciais, ou seja, o sistema é instável a pequenas perturbações. Para a região ordenada há uma quantidade diminuta de tipos celulares (atratores). A diversidade celular é pequena e cresce linearmente com N . Nessa região, o nosso resultado é similar ao encontrado no trabalho de Aldana [9], ou seja, crescimento linear do número de atratores diferentes e dos seus períodos com o número de elementos na rede. Por sua vez, o crescimento exponencial do número de atratores ou tipos celulares na região caótica, concorda com o previsto no trabalho recente de Kazumoto [32].

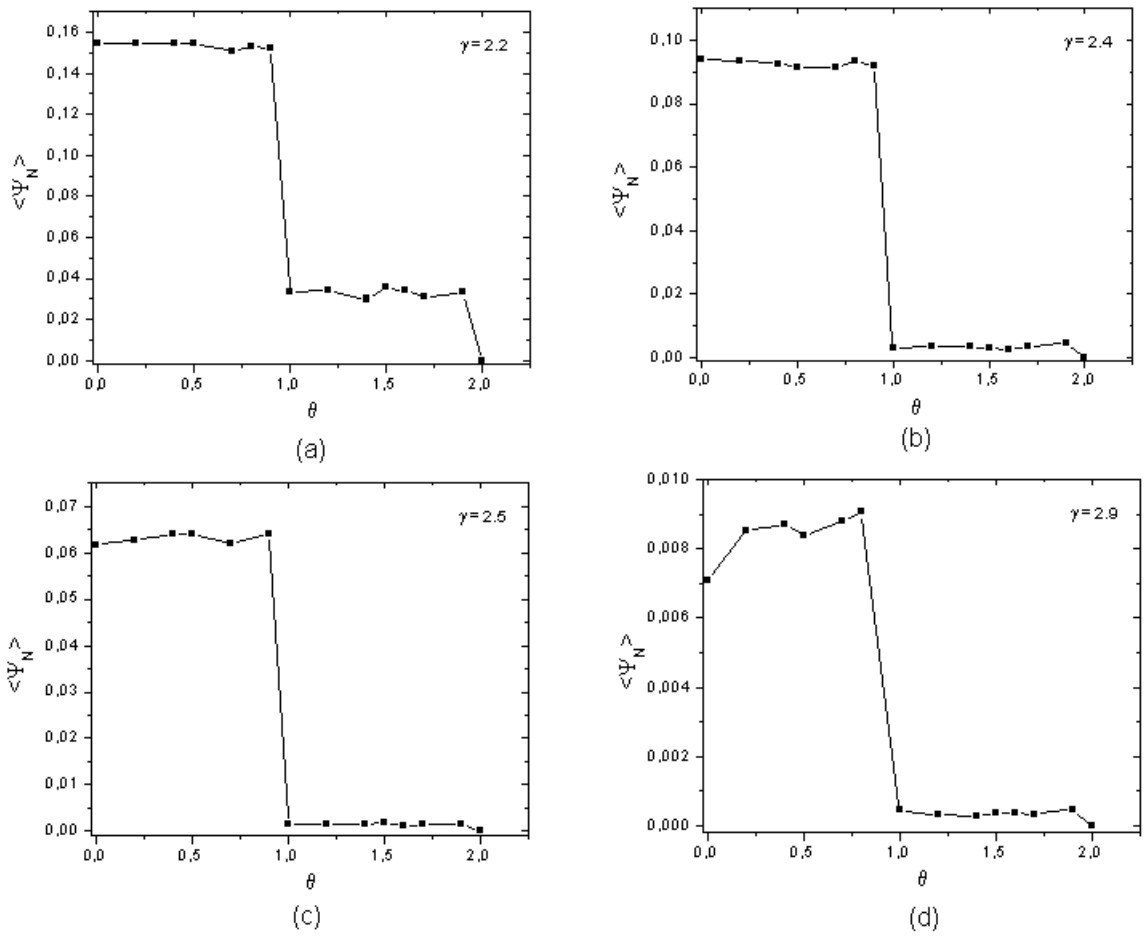


Figura 4.8: Dano final médio em função de θ para alguns valores de γ . (a) $\gamma = 2.2$ - esse expoente corresponde ao encontrado para a rede metabólica da bactéria *Escherichia coli* [6]. (b) $\gamma = 2.4$ - valor determinado para a rede de proteínas da levedura *Saccharomyces cerevisiae* [6]. (c) $\gamma = 2.5$. (d) $\gamma = 2.9$. Os dados correspondem a 200 redes diferentes com $N = 10000$ genes.

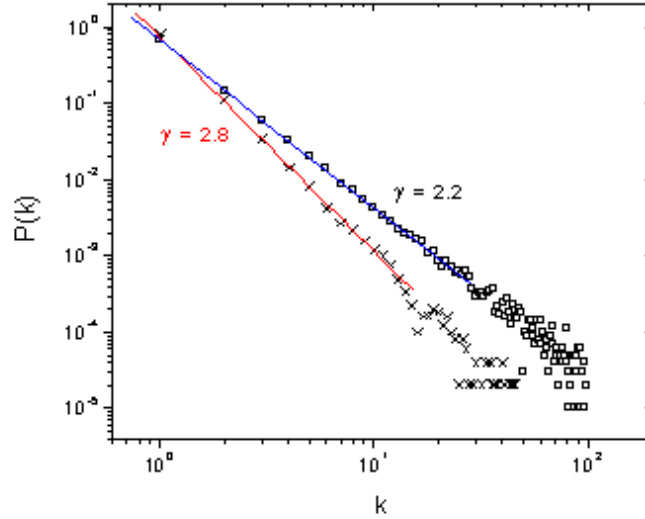


Figura 4.9: Distribuição de conectividades para redes livres de escalas com valores diferentes de γ , o expoente que controla a distribuição de grau de conectividade $P(K)$. Os dados correspondem a médias realizadas em redes com $N = 2000$ genes e 1000 amostras diferentes.

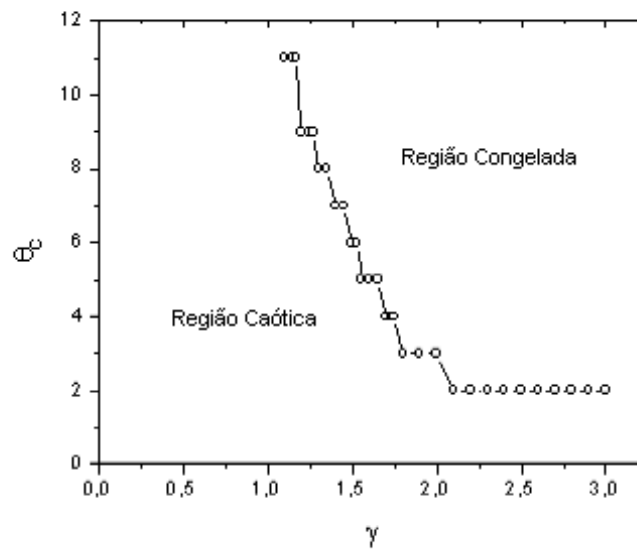


Figura 4.10: Diagrama de fases para o modelo de Sales-Martins-Stariolo para redes genéticas com topologia livre de escala. As fases caóticas e congeladas referem-se ao comportamento da distância de Hamming $\Psi(t)$ para $t \rightarrow \infty$; Na fase congelada $\Psi = 0$ enquanto que na fase caótica Ψ/N tende a um valor finito não nulo.

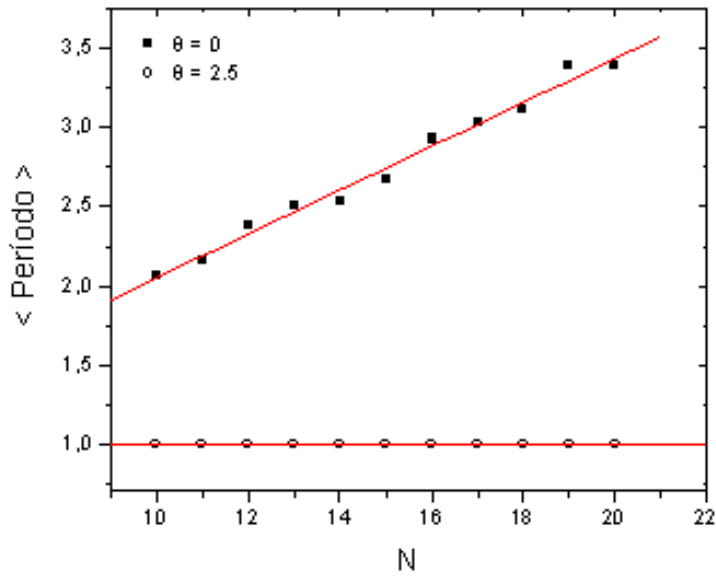


Figura 4.11: Período dos atratores no modelo de Sales-Martins-Stariolo em função do número de genes N na rede genética. As médias correspondem a 100 redes diferentes para o expoente $\gamma = 2.25$ e para os valores de $\theta = 0$ (fase caótica) e $\theta = 2.5$ (fase congelada).

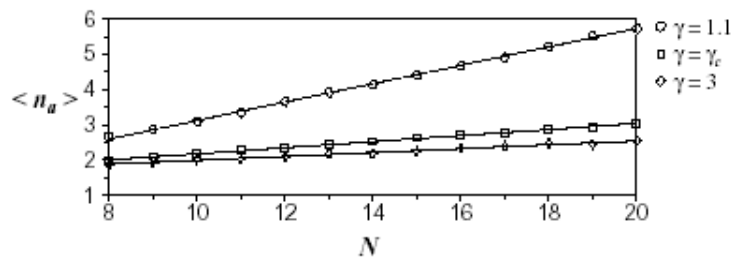


Figura 4.12: Número médio de atratores em função do tamanho da rede genética para valores de γ diferentes. Figura retirada da Ref. [9].

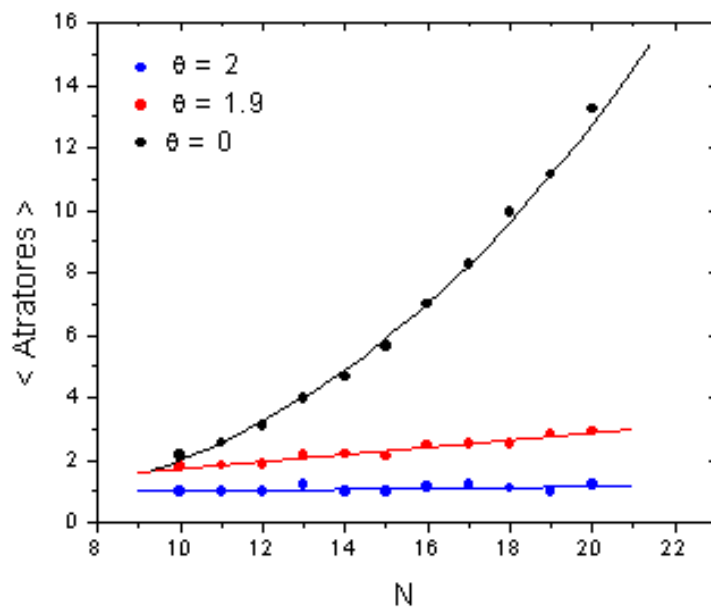


Figura 4.13: Número médio de atratores (células diferentes) no modelo de Sales-Martins-Stariolo como função do número de genes N . Na região caótica há uma maior diversidade celular. As médias correspondem a 1000 redes diferentes para o expoente $\gamma = 2.25$.

Capítulo 5

Conclusões

A rede de interação genética existente no núcleo celular de todas as células dos organismos vivos é um sistema altamente complexo. De fato, os genes regulam as atividades uns dos outros através de uma rede de interações moleculares entre proteínas. A transcrição é regulada por proteínas que interagem em seqüências de DNA, tal como elementos promotores, seqüências do consenso TATA, elementos estranhos, e assim por diante. As atividades de promotores, por sua vez, são controladas por fatores de transcrição, freqüentemente proteínas codificadas por outros genes, que se difundem no núcleo ou célula, ligam-se a esses sítios e regulam seu comportamento. Indo além do genoma, a tradução é regulada por uma rede de sinais, assim como também as atividades de uma gama de enzimas, cujos estados de fosforilação governam as atividades catalíticas e de ligação. O estado de fosforilação, por sua vez, é controlado por outras enzimas, as kinases e fosforilases, que são elas mesmas fosforiladas e desfosforiladas. O genoma e seu produto direto e indireto, em resumo, constituem uma rede complexa de interações moleculares [33]. O estudo dessa complexidade pode ser feito com base em modelos simples de controle de genes em redes com topologia

livre de escala. A simulação computacional da propagação de danos no modelo de autômatos celulares de Sales-Martins-Stariolo em redes livres de escalas para a regulação gênica demonstrou a existência de duas fases distintas: congelada e caótica. Um dano genético (mutação) tende a se propagar mais rapidamente na fase caótica, propiciando assim um maior número de células diferenciadas (atratores), número que cresce exponencialmente com o tamanho N da rede. A fase congelada, por sua vez, é mais rígida a mudanças em sua estrutura dinâmica e o número de atratores apresenta um crescimento linear com N . Verificamos que o número de conexões iniciais por gene m possibilita identificar em qual região o modelo encontra-se. Para a rede de Barabási-Albert ($\gamma = 3$) e ausência de canalização ($\theta = 0$), apenas para $m = 1$ a rede de genes está na região ordenada e para valores maiores ela está no regime caótico. O parâmetro θ , o limiar de ativação dos genes, força o sistema para a fase congelada (canalização) possibilitando um controle maior sobre o sistema. Em redes livres de escalas com o expoente da distribuição de conectividade variável, a fronteira entre as fases congelada e caótica é determinada por esse expoente γ e pelo limiar θ . Em particular o regime biológico (ordem com flexibilidade para mutações e seleção) surge na região $\gamma \geq 2.4$ e $1 \leq \theta < 2$. É importante observar que os resultados experimentais apontam para valores de $\gamma \sim 2.5$ para o mapa de interação entre proteínas e $\gamma \sim 2.3$ para a distribuição de conectividade no núcleo do metabolismo da *E. coli*.

Capítulo 6

Perspectivas Futuras

As redes com topologia livre de escala são, em primeira aproximação, bons modelos para o estudo de redes biológicas como, por exemplo, o mapa de regulação gênica. Entretanto, além da arquitetura livre de escala, as redes biológicas também apresentam modularidade. Esta modularidade é criada quando sítios organizam-se em grupos de forma hierárquica, enquanto mantém a topologia livre de escala. Nas redes em que há a presença de modularidade pode-se identificar facilmente grupos que são altamente conectados, nos quais poucos sítios ligam-se com outros sítios que se encontram em grupos distintos. Portanto, uma perspectiva futura para o presente trabalho é a simulação computacional de modelos definidos em redes modulares para o estudo das interações gênicas. O algoritmo para as redes hierárquicas já foi construído e um próximo passo é simular o modelo de Sales-Martins-Stariolo nessa rede. Outro aspecto a ser investigado, se ocorre ou não, é o comportamento coordenado, e em que faixa do espaço de parâmetro, de um sistema de ACs de Sales-Martins-Stariolo acoplados que pode ser usado para modelar a ontogenia e a diferenciação celular nos estágios iniciais do desenvolvimento embriológico. Este estudo seria uma

extensão natural daquele realizado em [34]. Finalmente, é importante estudar se uma dinâmica ordenada e reprodutível ocorre no modelo de Sales-Martins-Stariolo quando a atualização sincrônica da rede é perdida e cada elemento (gene) processa sinais que chegam em tempos que flutuam. Será que nesse caso o sistema tende a se auto-sincronizar? Provavelmente a resposta estará correlacionada com a topologia da rede subjacente. Isto porque o tempo de transição entre estados em respostas às flutuações nas concentrações de proteínas reguladoras varia com a distância entre os nós.

Referências Bibliográficas

- [1] VOET, D., VOET, J. G., PRATT, C. W.; *Fundamentos de Bioquímica. Artes Médicas*, (2002).
- [2] ALBERTS, B., JOHNSON, A., LEWIS, J., RAFF, M., ROBERTS, K., WALTER, P.; *Biologia Molecular da Célula. 4º ed., Artes Médicas*, (2006).
- [3] LEHNINGER, A. L., NELSON, D. L.; *Princípios de Bioquímica. 3º ed., Sarcier*, (2002).
- [4] VÊNICIO, R. Z. N., OLIVEIRA, G. B.; *Integração entre Biologia e Física: construindo redes de regulação gênica usando equações diferenciais. Rev. Bras. Ens. Bioq. Biol. Mol.* **1** (2006).
- [5] NEWMAN, M. E. J.; *The structure and function of complex networks. SIAM Review* **45**(2): 167-256 (2003).
- [6] ALBERT, R., BARABÁSI, A. L.; *Statistical Mechanics of Complex Networks. Rev. Mod. Phys.* **74**: 47-97 (2002).

- [7] ALMAAS, E., BARABÁSI, A. L.; *Power laws in biological networks. Capítulo do "Power laws, scale free networks and genome biology"*, E. Koonin ed., Landes Bioscience, (2005).
- [8] BARABÁSI, A. L., OLTVAI, Z. N.; *Network Biology: Understanding the cell's functional organization*. Nature Reviews Genetics **5**: 101-113 (2004).
- [9] ALDANA, M.; *Boolean dynamics of networks with scale-free topology*. Physica D **1**: 185 (2003).
- [10] JEONG, H., TOMBOR, B., ALBERT, R., OLTVAI, Z. N., BARABÁSI, A. L.; *The large-scale organization of metabolic networks*. Nature **407**: 651-654 (2000).
- [11] FARKAS, I., JEONG, H., VICSEK, T., BARABÁSI, A. L., OLTVAI, Z. N. *The topology of the transcription regulatory network in the yeast, Saccharomyces cerevisiae*. Physica A **318**: 601-612 (2003).
- [12] LUSCOMBE, N. M., BABU, M. M., YU, H., SNYDER, M., TEICHMANN, S. A., GERSTEIN, M.; *Genomic analysis of regulatory network dynamics reveals large topological changes*. Nature **431**: 308-312 (2004).
- [13] UETZ, P., ROTHBERG, J. M., et al.; *A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae*. Nature **403**: 623-627 (2000).
- [14] GIOT, L., ROTHBERG, J. M., et al.; *A Protein Interaction Map of Drosophila melanogaster*. Science **302**: 1727-36 (2003).

- [15] HARTWEEL, L. H., HOPFIELD, J. J., LEIBLER, S., MURRAY, A. W.; *A From molecular to modular cell biology*. Nature **402**: C47-52 (1999).
- [16] MERING, C. V., ZDOBNOV, E. M., TSOKA, S., CICCARELI, F. D., LEAL, J. B. P., OUZOUNIS, C. A., BORK, P.; *A Genome evolution reveals biochemical networks and functional modules*. Proc. Natl. Acad. Sci. USA **100**: 15428-15433 (2003).
- [17] COLIZZA, V., FLAMMINI, A., MARITAN, A., VESPIGNANI, A.; *Characterization and modeling of proteinprotein interaction networks*. Physica A **352**: 1-27 (2005).
- [18] RAVASZ, E., SOMERA, A. L., MONGRU, D. A., OLTVAI, Z. N., BARABÁSI, A. L.; *Hierarchical organization of modularity in metabolic networks*. Science **297**: 1551-1555 (2002).
- [19] RAVASZ, E., BARABÁSI, A. L.; *Hierarchical organization in complex networks*. Phy. Rev. E **67**026112 (2003).
- [20] TAKEMOTO, K., OOSAWA, C.; *Evolving networks by merging cliques*. Phys. Rev. E **71**046116 (2005).
- [21] EISENBERG, E., LEVANON, E. Y.; *Preferential Attachment in the Protein Network Evolution*. Phys. Rev. Lett. **91**(13) 138701 (2003).
- [22] BARABÁSI, A. L., ALBERT, R.; *Emergence of Scaling in Random Networks*. Science **286**: 509-512 (1999).

- [23] WOLFRAM, S.; *Statistical mechanics of cellular automata*. Rev. Mod. Phys. **55**: 601-644 (1983).
- [24] LEWIN, R.; *COMPLEXIDADE, A vida no limite do caos*. Rocco, (1994).
- [25] KAUFFMAN, S. A.; *The Origins of Order*. Oxford University Press, (1993.)
- [26] DERRIDA, B., POMEAU, Y.; *Random networks of automata: a simple annealed approximation*. Europhys. Lett. **1**: 45-49 (1986).
- DERRIDA, B., WEISBUCH, G.; *Evolution of overlaps between configurations in random Boolean networks*. J. Physique **47**: 1297-1303 (1987).
- [27] SALES, J. A., MARTINS, M. L., STARIOLO, D. A.; *Cellular automata model for gene networks*. Physical Review E **55**(3): 3262 (1997).
- [28] DERRIDA, B., STAUFFER, D.; *Phase transitions in two dimensional Kauffman cellular automata*. Europhys. Lett. **2**(10): 739-745 (1986).
- [29] SOCOLAR, J. E. S., KAUFFMAN, S. A.; *Scaling in Ordered and Critical Random Boolean Networks*. Phys. Rev. Lett. **90**(6) (2003).
- [30] ALDANA, M.; *Boolean Dynamics with Random Couplings*. Perspectives and Problems in Nonlinear Science, 23-89 (2003).
- [31] KAUFFMAN, S. A.; *A proposal for using the ensemble approach to understand genetic regulatory networks*. J. Theor. Biol. **230**(4): 581-590 (2004).
- [32] IGUCHI, K., KINOSHITA, S.; *cond-mat/0510430* (2006).

- [33] KAUFFMAN, S. A.; *O que é vida? Schrödinger estava certo? In: O que é vida? 50 anos depois: especulações sobre o futuro da Biologia.* M.P. Murphy L.A.J. O'Neill (Eds.). São Paulo, UNESP/Cambridger, 101-135, (1997).
- [34] SILVA, H. S., MARTINS, M. L.; *A cellular automata model for cell differentiation.* Physica A **322**, 555-566 (2003).