

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA

**REDES EM BIOLOGIA: INTRODUÇÃO ÀS REDES COMPLEXAS,
ESTUDO DOS ASPECTOS ESTRUTURAIS E DINÂMICOS DO CICLO
CELULAR E DOS RITMOS CIRCADIANOS**

Dissertação apresentada ao Curso de Pós-Graduação
em Física da Universidade Federal de Minas Gerais
como requisito parcial para a obtenção do grau de
Mestre em Física.

LUCAS LAGES WARDIL

Orientado por Jaferson K. L. da Silva

Co-orientado por Alcides Castro e Silva

Belo Horizonte
11 de fevereiro de 2008

Abstract

In the post-genomic era, there is a lot of data about intricate networks of interaction among the units inside a cell. Setting general and fundamental principles from this data is now a challenge and useful models have been proposed to capture such principles. Usually, the data and the processes can be represented by complex networks. We present an introduction to biological complex networks, in which we have seen that lots of biological processes can be represented by such networks. But if the biological processes are very complex are the models necessarily complex? We show, for the circadian clock of *Drosophila melanogaster* and mammals, that if the fundamental principle is stability a very simple model can capture it. Not just the stability is indeed verified but other biological features, like the modification of the circadian clock by means of a transgenic experiment, are also studied. We also studied the cell cycle of the yeast *S. cerevisiae*, in which not just stability, but robustness, can be captured by the same simple model, for which we have shown that the inhomogeneous version is closer to the *in vivo* systems.

Resumo

Na era pós-genômica, há uma quantidade imensa de dados relacionados às intrincadas redes de interações entre as unidades que compõem uma célula. Formular princípios gerais e fundamentais a partir desses dados é hoje um desafio, e vários modelos foram propostos para capturar tais princípios. Em geral, os dados e os processos podem ser representados por redes complexas. Nós apresentamos uma introdução às redes complexas em biologia, onde verificamos que muitos processos biológicos podem ser representados por tais redes. Mas, se o processo biológico é complexo, os modelos precisam ser necessariamente complexos? Nós mostramos, para o oscilador circadiano da *Drosófila melanogaster* e dos mamíferos, que, se o princípio fundamental é a estabilidade, então um modelo bastante simples é suficiente. Não apenas a estabilidade é de fato verificada mas ainda outras características biológicas, como, por exemplo, a modificação do oscilador circadiano através de um experimento transgênico, são também estudadas. Nós também estudamos o ciclo celular da levedura *S. cerevisiae* em que não só a estabilidade, mas também a robustez podem ser capturadas através do mesmo modelo, para o qual mostramos que a versão não homogênea é mais coerente com o sistema *in vivo*.

Dedico aos colaboradores anônimos do progresso social que, incansavelmente, contribuem com suas ínfimas parcelas para o avanço do conhecimento científico sem que, na maioria das vezes, vejam a consequência direta de seus esforços.

Agradecimentos

Agradeço aos meus pais Ricardo e Rosana e aos meus queridos irmãos, Mariana, Daniel, Cristiano e Fernanda. A essas pessoas as palavras são poucas, mas o sentimento de amor por vocês é imenso!

Agradeço a minha avó, Bete, pela revisão gramatical dessa dissertação.

Agradeço à Sylvie por ter-me conduzido ao meu orientador de iniciação científica, Mário Jorge Dias Carneiro, e ao meu atual orientador, Jafferson.

Agradeço ao Mário Jorge pela condução de meus primeiros passos na ciência e pelo incrível exemplo de humildade e sensatez. Muito obrigado!

Agradeço ao Jafferson, não só pela orientação nesse mestrado, mas por ter-me contagiado com sua alegria inestancável em fazer física. Você não só orienta a pesquisa, mas cuida do aluno com um envolvimento afetivo incrível.

Agradeço ao Alcides pela sugestão dos primeiros problemas de pesquisa.

Agradeço ao amigo Henrique Di Lorenzo. Devido à sua influência, reencantei-me com o estudo da física e aprendi a pensar como um físico. Sua personalidade singular foi o exemplo de que razão e simplicidade podem caminhar juntamente. Saudades!

Agradeço aos trabalhadores da Fundação Espírita Cárita, pois é com esses trabalhadores que eu concretizo alguns dos meus ideais visando a algumas das minhas metas de Espírito imortal que sou.

Agradeço a Jesus pela oportunidade de desenvolver nessa reencarnação dois aspectos simultaneamente: o religioso e o científico.

A gratidão a Deus não será feita por palavras, mas pela forma de trabalhar: trabalharei incansavelmente com uma dedicação amorosa à pesquisa, ainda que a preço de renúncia e sacrifício!

Sumário

1	Introdução	1
2	Redes	4
2.1	Conceitos elementares	4
2.2	Mecânica estatística de redes	7
2.3	Redes sem-escala	8
2.3.1	A propriedade “sem-escala”	8
2.3.2	Outras redes sem-escala	9
2.3.3	Propriedades notáveis	10
3	Redes biológicas	12
3.1	Rede de interação entre proteínas	13
3.2	Redes metabólicas	14
3.3	Redes de transcrição genética	15
3.4	Redes de tradução de sinais	16
4	Redes no equilíbrio	18
4.1	Redes aleatórias clássicas	19
4.2	Construção de redes em equilíbrio	20
4.2.1	Condensação	23
4.3	Redes “ <i>small-world</i> ”	24
5	Redes fora do equilíbrio	28
5.1	Crescimento aleatório	28
5.2	Crescimento de redes sem-escala através de ligações preferenciais	30
5.2.1	Um modelo com ligações preferenciais	31
5.2.2	Caso sublinear	32
5.2.3	Caso linear	32
5.2.4	Caso superlinear	33
5.2.5	Correlações entre as conectividades dos nós	33
5.3	Organização hierárquica	34
6	Das redes aos processos fisiológicos	37
6.1	Sistemas dinâmicos	37

6.1.1	Sistemas dinâmicos contínuos	40
6.2	Informação biológica	41
6.3	Controle da expressão gênica	43
6.3.1	Controle transcricional	43
6.3.2	Controles pós-transcricionais	44
6.4	Redes regulatórias e sistemas dinâmicos	45
6.5	Da biologia celular molecular para a modular	48
7	Ciclo celular	50
7.1	Panorama geral do ciclo celular	50
7.2	Sistema de controle do ciclo celular da <i>S. cerevisiae</i>	51
7.2.1	Rede regulatória	52
7.2.2	Descrição detalhada das interações de ativação e inibição	53
7.2.3	O modelo homogêneo	55
7.3	A robustez da rede	57
7.4	Intermissão	59
7.5	Aspectos do modelo homogêneo	59
7.5.1	Ligações duvidosas e circuitos ausentes	59
7.5.2	Variações no tempo de degradação implícita	60
7.6	O modelo não-homogêneo	61
7.6.1	Resultados e análises	63
8	Ciclo circadiano	65
8.1	Panorama geral	66
8.2	O oscilador circadiano da <i>D. melanogaster</i>	66
8.2.1	A rede regulatória	66
8.2.2	Descrição das interações de ativação e inibição	67
8.2.3	O modelo	68
8.3	A estabilidade e a robustez do oscilador circadiano da <i>D. melanogaster</i>	70
8.4	O ciclo circadiano dos mamíferos	73
9	Conclusão	75
	Referências Bibliográficas	76

Lista de Figuras

2.1	Representação gráfica de um grafo não direcionado (esquerda) e de um grafo direcionado (direita). Os círculos representam os vértices e os segmentos de reta representam as ligações. Os números referem-se aos nós.	4
2.2	Distância entre os nós. A distância entre os nós 1 e 9 é obtida percorrendo-se o menor caminho que liga ambos os vértices: {1,3,4,5,9}. Portanto a distância é 4.	6
2.3	O nó 1 é um hub da rede. Na ausência do hub, i.e., na ausência das ligações tracejadas, a distância entre os nós 2 e 7 seria 5. O hub cria um atalho reduzindo a distância para 2. . . .	11
3.1	Rede de interação de proteínas da <i>C. elegans</i> . A caixa em destaque apresenta uma parte da rede em detalhes. As interações foram obtidas a partir de testes de interação física e a partir de inferências similares às descritas na seção 3.1. [1]	14
3.2	Propriedades topológicas da rede de interação de proteínas da <i>S. Cerevisiae</i> obtidas a partir de quatro bancos de dados. (a) Distribuição de conectividade. A linha sólida representa a distribuição $P(k) \propto k^{-2.5}$. (b) Coeficiente de aglomeração. A linha sólida representa a função $C(k) = \beta/k^2$ [1]	15
3.3	Componente conexa da rede metabólica da <i>E. Coli</i> que contém a glicose. O nó maior na região central representa a glicose e o nó maior na região mais à esquerda representa o nó mais distante da glicose. [2]	16
3.4	Rede de reguladores da transcrição da <i>S. Cerevisiae</i> . Os nós são os fatores de transcrição e uma ligação do tipo $A \rightarrow B$ significa que o fator A atua na transcrição do fator B . As seis figuras representam os mesmos nós, mas cada figura separadamente contém as ligações responsáveis pela coordenação de algum processo específico. A figura no canto inferior contém as interações de todos os processos. [3]	17
4.1	Ilustração de um grafo do tipo árvore. Note a ausência de caminhos fechados.	21
4.2	Condensação em uma rede de equilíbrio para $\bar{k} > k_c$. A distribuição é reflexo de duas contribuições: uma coincidente com a distribuição no ponto crítico, e a outra devido à condensação das ligações em torno de um conjunto pequeno de vértices com conectividade em torno de um valor Q	24
4.3	Rede regular de dimensão 1 com condições periódicas de fronteira e $z = 4$	25
4.4	Estrutura local de um nó. As linhas tracejadas delimitam os ramos esquerdo e direito.	25
4.5	Distância média entre dois nós e coeficiente de aglomeração, normalizados por seus valores em $p = 0$, versus p . Esse gráfico caracteriza a transição entre redes regulares e aleatórias [4]. . . .	26

5.1	Crescimento aleatório. No tempo $t = 2$ inicia-se com os nós 1 e 2. No próximo passo, o nó é acrescentado e ligado aleatoriamente a um dos nós já existentes, que nesse figura é o nó 1. O processo se repete até que a rede tenha N nós e N ligações.	29
5.2	Mecanismo de duplicação, mutação e divergência. O gene a possui uma cópia aa e ambos produzem a mesma proteína A . Se uma das cópias sofre mutação, haverá dois genes, a e aam , e poderá haver uma divergência nas funções executadas pelas proteínas, resultando em duas proteínas distintas, A e Am . Se o resultado da mutação não for radical, a nova proteína Am permanece ligada à proteína B	30
5.3	Evolução temporal do crescimento da rede determinística. A cada passo uma ligação dá origem a um nó que se liga aos extremos dessa ligação.	35
5.4	Rede hierárquica organizada modularmente.	36
6.1	Exemplo de uma rede à qual pode-se vincular um sistema dinâmico. As ligações são direcionadas.	39
6.2	Órbitas no espaço dos vetores de estado. Cada nó representa uma configuração $(\sigma_1, \sigma_2, \sigma_3)$. Os atratores da dinâmica são os nós $(0, 0, 0)$ e $(1, 1, 1)$	40
6.3	Esquema da representação da informação contida no DNA em termos de uma rede. Na primeira etapa traduz-se o gene em proteína. Na segunda etapa, a especificidade das proteínas determina as ligações químicas entre elas que são representadas na rede de interação proteína-proteína. A segunda etapa é realizada pelos métodos descritos na seção 3.1.	42
6.4	Representação do sistema de controle transcricional. Se o gene A não produzir a proteína A, o gene B permanece inativo. Se o gene A produzir a proteína A, esta atua como fator de ativação na transcrição do gene B e a proteína B é produzida. Da mesma maneira que o gene B, o gene A é ativado ou desativado por algum outro fator.	44
6.5	Esquema do sistema de controle da expressão gênica. A etapa que compreende a transcrição dos genes pode ser controlada pelo controle transcricional. Todas as demais etapas subsequentes à transcrição são controladas pelos controles pós-transcricionais.	45
6.6	Exemplo de um pequeno circuito regulatório “ <i>feed-forward loop</i> ”.	46
6.7	Exemplo de um pequeno circuito regulatório “ <i>mult-input</i> ”.	46
6.8	Motifs presentes na rede regulatória da <i>S. cerevisiae</i> [3].	47
7.1	Etapas do ciclo celular.	51
7.2	Rede regulatória do ciclo celular apresentada em [5]. Os nós estão identificados por um número e pela proteína. As interações de ativação estão representadas pelas setas contínuas e as de inibição, pelas setas tracejadas. Um seta dupla significa que ambos os nós atuam como ativadores ou inibidores recíprocos.	53
7.3	Órbitas da maior bacia de atração. O ponto mais inferior na figura corresponde ao estado estacionário 272. [5]	59
8.1	Oscilador circadiano da <i>D. melanogaster</i> . As ligações com setas representam interações de ativação e as ligações com uma barra em seu término representam interações de inibição. As ligações tracejadas indicam a parte do oscilador que não foi considerada no estudo do sistema dinâmico, mas que estão presentes de maneira implícita na dinâmica. O contorno tracejado identifica os dois circuitos interconectados via CYC/CLK	67

8.2	Concentrações de algumas proteínas e RNA em função do tempo para um período do ciclo circadiano [6].	69
8.3	Órbitas no espaço de estados. O ciclo, que é o único atrator, está representado no centro da figura com os nós e setas maiores.	71
8.4	Rede regulatória do oscilador circadiano dos mamíferos. A linha tracejada separa os dois osciladores entrelaçados.	73

Lista de Tabelas

2.1	Algumas grandezas características das redes de co-autorias da Medline e SPIRES. N é a quantidade de nós. \bar{k} é a conectividade média. C é o coeficiente de aglomeração médio. \bar{l} é a distância média entre os nós e l_{max} é a maior distância entre dois nós.	10
2.2	Medidas realizadas em algumas redes reais. N representa a quantidade de nós e L a quantidade de ligações. γ é o coeficiente da lei de potência que representa a distribuição de conectividade. C/C_r (\bar{l}/\bar{l}_r) representa a razão entre o coeficiente de aglomeração médio (distância média) da rede real e o coeficiente de aglomeração (distância média) de uma rede aleatória clássica com a mesma quantidade de nós e ligações (vide capítulo 4 para a definição da rede aleatória clássica).	10
7.1	Tamanho BS das bacias de atração. Cada linha representa os vetores de estado que são atratores da dinâmica.	57
7.2	Órbita da configuração inicial correspondente ao estado estacionário perturbado. A coluna tempo corresponde aos passos temporais; as colunas identificadas pelos nós correspondem à evolução temporal dos mesmos; a coluna fase corresponde às fases do ciclo celular; a coluna notação corresponde à notação do vetor de estado de acordo com 7.15.	58
8.1	Órbita correspondente ao ciclo de período 12 que é o único atrator da dinâmica.	70

Capítulo 1

Introdução

Durante muito tempo, os biólogos estudaram como as partes de uma célula funcionam: a estrutura do DNA, a das proteínas, os processos bioquímicos etc.. Esses esforços geraram uma caracterização riquíssima sobre a célula. Complementar a esses estudos, iniciou-se a investigação de como as unidades componentes da célula interagem a fim de executar as atividades celulares. A pergunta que surgia era do tipo: como as células se comunicam para executar uma tarefa em comum? Como os genes interagem entre si a fim de garantirem um programa de controle rigoroso de certa atividade celular? [7] Essa abordagem, que tem por objetivo entender como as partes interagem entre si na execução de uma função, foi chamada biologia sistêmica [7]. Mas, de repente, a abordagem sistêmica viu-se imersa num mar de dados necessitando obter informações desses dados. Surgia a era pós-genômica [8], período que se seguiu ao seqüenciamento do genoma humano. Os genomas de várias espécies foram seqüenciados e uma avalanche de genes e proteínas foram sendo catalogados com as respectivas interações: proteína A interage com as proteínas B_1, B_2, \dots , o gene α inibe os genes β_1, β_2, \dots . Mas, qual o significado desses dados?

Parte das informações provenientes da abordagem sistêmica são bem representadas através do objeto matemático comumente conhecido como rede. Uma rede é um conjunto de elementos (os nós da rede) que possuem ligações entre si. Pictoricamente pode ser visualizada como um conjunto de pontos ligados entre si por linhas. Dessa forma, os dados obtidos pelos projetos genomas (e seus derivados, proteoma, interactoma etc.) mostraram-se aptos a serem representados por redes complexas, em que a caracterização das interações das unidades básicas de uma célula, como os genes, proteínas, metabólitos etc., é evidenciada em seu aspecto global. Nesse caso, os genes, proteínas, metabólito etc. são representados pelos nós e as interações pelas ligações na rede.

Como um processo biológico não é executado somente por um elemento, mas pela interação de múltiplas unidades que, numa interação complexa, executam as atividades mantenedoras da vida [8], as redes complexas são uma representação dos processos biológicos como um todo. A questão que surgia era como as funções biológicas emergem da rede complexa de interações? É sabido que cada função biológica é executada por um conjunto de elementos (genes, proteínas etc.) que se configuram como um módulo funcional, isto é, configuram-se como uma entidade discreta cuja função é separável de outros módulos [9]. Uma questão que surgiu era se os módulos funcionais correspondiam aos módulos estruturais (conjuntos de nós altamente conectados entre si mas pouco conectados com outros nós na rede). A estrutura da rede indicou a presença de uma estrutura hierárquica de módulos estruturais. Estudos iniciais parecem indicar que esses módulos exercem funções específicas, embora não se esteja totalmente certo quanto à

significância de tal correlação módulo funcional e módulo estrutural [10]. Contudo, é certo que determinados conjuntos de nós possuem funções específicas e executam atividades biológicas específicas, apesar de haver nós que possuem mais de uma função [11, 12]. Por exemplo, é bem definido que (i) o conjunto de proteínas que controlam a divisão de uma célula pode ser representado por uma rede de proteínas [5]; (ii) o conjunto de genes que regulam o desenvolvimento de um embrião pode ser representado em uma rede regulatória de transcrição [13]; (iii) as reações metabólicas que executam a síntese e a degradação da matéria-prima necessária para a manutenção de um sistema vivo podem ser representadas em uma rede metabólica [14].

A existência de nós que exercem várias funções, participando de vários módulos funcionais, foi observada na estrutura da rede como sendo nós altamente conectados, isto é, nós com muito mais ligações do que a média [11]. A esses nós deu-se o nome “hubs” e às redes que possuem esses nós deu-se o nome redes sem-escala [15]. O fato notável é que essas redes se mostraram ubíquas em a natureza e várias redes biológicas foram consideradas sem-escala. Essas redes se mostram tão abundantes porque forças evolutivas estão vinculadas à formação da estrutura de redes sem-escala [16, 17]. Nessa dissertação, apesar de apresentarmos uma revisão da teoria de redes, nós focaremos o estudo apenas nos conjuntos de nós que realizam uma determinada função e não no aspecto estrutural de uma rede sem-escala. Em outras palavras, nós não trabalharemos com toda a estrutura complexa da rede, mas apenas como uma parte dela, os módulos funcionais que executam determinadas tarefas específicas.

É interessante notar que uma proteína sozinha é considerada um objeto não-vivo. Mas a organização de várias unidades não-vivas, em alguns casos, manifesta a propriedade de sistemas vivos. Uma rede parece, portanto, um objeto capaz de capturar essencialmente essa propriedade de que vivo, só o é o organismo inteiro.

Contudo, as redes biológicas não contêm a informação da evolução temporal das atividades biológicas. Para isso, é necessário vincular à rede um conjunto de regras que determinem a evolução temporal, isto é, é necessário vincular um sistema dinâmico como sendo o modelo da evolução temporal. O estudo dos sistemas dinâmicos fornecerá o conceito para formalizar algumas noções de significado biológico, como robustez, estabilidade, perturbação etc.

É, portanto, através dessas duas ferramentas formais, as redes e os sistemas dinâmicos, que tentaremos compreender, sob o aspecto sistêmico, duas tarefas específicas executadas por praticamente todos os organismos vivos: o ciclo celular e os ritmos circadianos. O ciclo celular é a atividade em que a célula gera cópias de si mesma duplicando-se. Os ritmos circadianos são processos fisiológicos que possuem um ritmo vinculado ao ritmo dia-noite, sendo que esses ritmos são controlados por processos regulatórios que determinam as oscilações circadianas.

No primeiro capítulo, oferecemos uma introdução aos conceitos básicos da teoria de redes. Na seção 2.3.2 desse primeiro capítulo ilustramos algumas áreas em que o conceito de rede é utilizado e que foram campo das primeiras pesquisas de redes complexas. Logo em seguida, no capítulo 3, elencamos alguns exemplos de redes biológicas que fornecerão ao leitor uma noção melhor do objeto com que estamos trabalhando. Nos capítulos 4 e 5, abordamos a teoria de redes usando ferramentas típicas da mecânica estatística, tratando as redes em equilíbrio e fora de equilíbrio. No capítulo 6, apresentamos a conexão entre a teoria de redes e os processos fisiológicos, abordando nesse capítulo a teoria de sistemas dinâmicos e a teoria do controle da expressão gênica. Os capítulos 7 e 8 se dedicam ao estudo de dois processos biológicos governados por redes regulatórias: o ciclo celular e o ritmo circadiano.

Como os capítulos iniciais são dedicados a uma revisão da teoria de redes e de alguns tópicos em biologia, e o capítulo sobre o ciclo celular inicia com a apresentação de resultados obtidos por Fanting Li et al. [5], dividimos a dissertação em duas partes: revisão teórica e apresentação de resultados obtidos por nós. A seção 7.4, intitulada “Intermissão”, demarca estas duas partes. Entre os nossos resultados, encontram-se algumas análises finais sobre o ciclo celular e todo o trabalho sobre ciclo circadiano. Em particular estudaremos a robustez e estabilidade dessas atividades.

Apesar do entendimento de “o que é a vida” ainda continuar a atormentar e a motivar o esforço de muitas gerações e mesmo tendo consciência de que os nossos trabalhos restringem-se à resolução de problemas periféricos, no que diz respeito a esse entendimento, aprendemos que, “embora tenhamos de enfrentar e compreender os grandes e abrangentes temas do nosso universo, o melhor é abordá-los através das pequenas curiosidades que prendem a nossa atenção—todos aqueles belos seixos na praia do conhecimento. O oceano da verdade banha os seixos a cada onda, e eles retinem produzindo o som mais maravilhoso que existe”, como diria o biólogo Stephen Jay Gould, autor de “Vida Maravilhosa” [18].

Capítulo 2

Redes

2.1 Conceitos elementares

O termo rede é utilizado em várias áreas do conhecimento sendo que, em cada área, o conceito rede refere-se a contextos distintos. Entretanto, abstração feita à diversidade contextual, uma rede é um objeto matemático conhecido como grafo [19].

Grafo: um grafo G é um par ordenado de conjuntos disjuntos (V, L) tal que L é um subconjunto do conjunto de pares não ordenados de V . O conjunto V é chamado de conjunto dos vértices e o conjunto L de conjunto das ligações. Diz-se que um elemento (v_1, v_2) de L liga os vértices v_1 e v_2 e que v_1 e v_2 são vértices adjacentes. Duas ligações são adjacentes se elas compartilham um mesmo vértice.

O conjunto L de ligações pode ser considerado um conjunto de pares ordenados quando então se têm um grafo direcionado. A figura 2.1 ilustra o conceito de grafo.

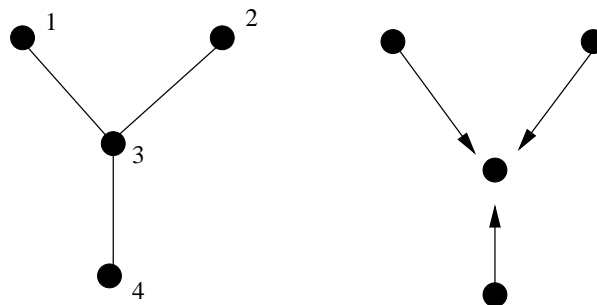


Figura 2.1: Representação gráfica de um grafo não direcionado (esquerda) e de um grafo direcionado (direita). Os círculos representam os vértices e os segmentos de reta representam as ligações. Os números referem-se aos nós.

Primeiramente, entremos em consenso em relação à terminologia: no presente trabalho, não mais nos referiremos ao termo grafo, mas somente ao termo rede, utilizando-os como sinônimos. Da mesma forma, usaremos os termos nó e vértice indistintamente. Outro termo que usaremos indiscriminadamente é o termo topologia. Ao nos referirmos à topologia da rede, apenas gostaríamos de nos referir à estrutura

de conexões entre os nós, sem nenhum outro compromisso maior com a teoria matemática de topologia. Feitas essas primeiras considerações, procedamos ao estudo de algumas definições básicas [20].

A informação topológica de uma rede pode ser completamente representada em uma matriz conhecida como matriz de adjacência.

Matriz de adjacência: é a matriz que contém a informação sobre as ligações entre os pares de nós (i, j) .

Os elementos $a_{i,j}$ assumem os seguintes valores: $a_{i,j} = 1$ se o par (i, j) pertence ao conjunto L das ligações e $a_{i,j} = 0$ caso contrário. Se a rede contém N nós, a matriz é $N \times N$.

Note que não estamos considerando múltiplas ligações, isto é, um mesmo par de vértices com mais de uma ligação entre eles. É evidente que a matriz é simétrica, $a_{i,j} = a_{j,i}$, se a rede é não direcionada. No caso de um grafo direcionado, se houver uma ligação do vértice i para o vértice j , então $a_{i,j} = 1$. A matriz de adjacência do grafo não direcionado da figura 2.1 é

$$\begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \quad (2.1)$$

Várias medidas podem ser computadas a partir da matriz de adjacência. Uma medida importante é o número de conexões de um vértice. No caso de uma rede não direcionada, o número de conexões, ou conectividade ou ainda grau do vértice é a quantidade de ligações de que o vértice participa. A partir da matriz de adjacência, a conectividade k_i i é

$$k_i = \sum_{j=1}^N a_{i,j}.$$

Assim, considerando o grafo não direcionado da figura 2.1, tem-se que $k_1 = k_2 = k_4 = 1$ e $k_3 = 3$. No caso de uma rede direcionada, a quantidade de ligações que incidem em um vértice ou que partem do mesmo quantifica medidas de conectividade distintas, $k_{in,i}$ e $k_{out,i}$, respectivamente.

A conectividade da rede pode ser caracterizada por uma medida global, i.e., referente à rede, e não a um nó: a distribuição de conectividade.

Distribuição de conectividade: é a probabilidade de se ter um vértice com uma dada conectividade.

Se o número de vértices N for grande o suficiente, então a distribuição de conectividade total da rede será

$$P(k) = \frac{1}{N} \sum_{s=1}^N \delta_{k_s,k}, \quad (2.2)$$

em que $\delta_{k_s,k}$ é o delta de Kronecker ($\delta_{i,j} = 1$ se $i = j$ e $\delta_{i,j} = 0$ se $i \neq j$). Novamente é possível definir a distribuição de conectividade em uma rede direcionada de maneira a distinguir as ligações que chegam a um vértice ou partem do mesmo. Seja k_{in} a quantidade de ligações que apontam na direção de um dado vértice e k_{out} a quantidade de ligações que partem desse mesmo vértice. Define-se $P(k_{in})$ e $P(k_{out})$ de maneira análoga à distribuição total e esta se relaciona às anteriores através de $P(k) = P(k_{in}) + P(k_{out})$.

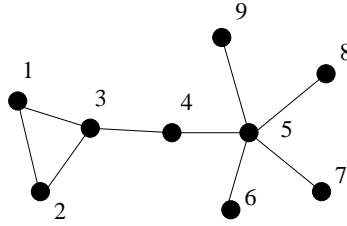


Figura 2.2: Distância entre os nós. A distância entre os nós 1 e 9 é obtida percorrendo-se o menor caminho que liga ambos os vértices: $\{1,3,4,5,9\}$. Portanto a distância é 4.

Como será apresentada mais adiante, a distribuição de conectividade classifica as redes que ocorrem na natureza em classes distintas.

A distribuição de conectividade fornece uma medida global. Uma, dentre as possíveis caracterizações locais, é obtida via o coeficiente de aglomeração (em inglês “clustering coefficient”).

Coefficiente de aglomeração: caracteriza a densidade de conexões em torno de um vértice.

Consideremos uma rede não direcionada e fixemos um nó i . Seja z_i a quantidade de vértices adjacentes a esse nó e y_i a quantidade de ligações existentes entre esses vértices adjacentes. Se todos os vizinhos estivessem ligados, a quantidade de ligações L_{max} seria a combinação de z_i nós dois a dois,

$$L_{max} = \binom{z_i}{2} = z(z-1)/2,$$

pois não estamos considerando ligações de um vértice consigo mesmo. Como nem todas as ligações entre os vizinhos estão presentes, a densidade de ligações em torno de um vértice i será

$$C_i = \frac{2y_i}{z_i(z_i-1)}.$$

Se a rede for suficientemente grande, é possível definir a distribuição dos coeficientes de aglomeração. O valor médio do coeficiente é, portanto, uma caracterização de quão aglomerados são os ambientes locais de uma rede.

Assim como é possível definir um caminho em uma superfície contínua, pode-se definir um caminho no espaço topológico discreto de uma rede.

Caminho: uma seqüência de vértices $\{v_1, \dots, v_i, v_{i+1}, \dots, v_t\}$ tal que v_i e v_{i+1} são vértices adjacentes e $v_i \neq v_{i+1}$ define um caminho na rede.

O comprimento de um caminho é definido como a quantidade de ligações que separam o vértice inicial do vértice final. Uma rede é conexa se, para todo par de vértices, existe um caminho ligando-os. Caso a rede não seja conexa, ela é sempre dividida em suas componentes conexas. Em uma mesma componente conexa é possível definir a distância entre dois vértices:

Distância: a distância entre dois vértices, em uma mesma componente conexa da rede, é definida como o comprimento do menor caminho que liga ambos os vértices.

Note que, se a rede for direcionada, é possível que haja um caminho ligando um vértice a outro e que não haja o caminho inverso. É possível atribuir comprimentos a cada ligação de maneira que o comprimento de um caminho seja a soma do comprimento das ligações que o compõe. Novamente, se a rede for grande o suficiente, é possível definir uma distribuição de distâncias $P(l)$ e uma distância média $\bar{l} = \sum_l lP(l)$. A figura 2.2 ilustra o conceito de distância.

Várias medidas podem ser definidas a fim de se quantificar os aspectos de interesse. Outra medida interessante é centralidade de um vértice. Seja $B(i, j) \geq 0$ a quantidade de caminhos mais curtos entre os vértices i e j e seja $B(i, m, j)$ a restrição da quantidade anterior aos caminhos que passam pelo vértice m . Então $B(i, m, j)/B(i, j)$ indica a participação do vértice m nos caminhos mais curtos entre i e j . A participação do vértice m em relação a todos os possíveis pares de vértices na rede é quantificado pela centralidade:

$$\sigma(m) = \sum_{i \neq j \neq m} \frac{B(i, m, j)}{B(i, j)}, \quad (2.3)$$

em que a soma é entre todos os possíveis pares de vértices para os quais exista um caminho ligando-os e que passe por m .

Centralidade: a centralidade $\sigma(m)$ de um vértice m é o número total de caminhos mais curtos entre todos os pares de vértices que passam pelo vértice m .

A teoria de grafos desenvolve uma série de resultados rigorosos a partir de vários conceitos que exploram outras caracterizações de um grafo. Para alcançar os nossos fins, apenas os conceitos elementares acima são suficientes.

2.2 Mecânica estatística de redes

Usualmente define-se mecânica estatística como sendo uma teoria que fornece a ponte entre a física microscópica e a macroscópica [21]. A dinâmica de partículas individuais regidas por uma hamiltoniana está relacionada à informação de um sistema de muitas partículas via o formalismo da mecânica estatística. No caso clássico, o formalismo é desenvolvido no espaço de fase, e grandezas macroscópicas podem ser obtidas a partir de médias realizadas no espaço de fase.

Mas no caso em que o sistema físico complexo é representado por redes, por exemplo, uma rede de interação de proteínas em que cada nó representa uma proteína e cada ligação corresponde à possibilidade de duas proteínas se ligarem formando um complexo (vide capítulo 3), é totalmente inviável representar as proteínas como partículas sujeitas a uma interação definida por uma hamiltoniana. É evidente que as proteínas se ligam via interações eletromagnéticas ou outras, mas a extrema complexidade dessas interações inviabiliza o programa de modelagem através de uma dinâmica hamiltoniana. Não obstante essa impossibilidade, é fato que das interações complexas surgem grandezas macroscópicas que caracterizam tais sistemas.

Como então definir uma mecânica estatística que possa descrever tais sistemas biológicos e estudar sua evolução temporal? Certos tratamentos da mecânica estatística parecem aplicáveis ao estudo das redes biológicas, como o estudo de um sistema via equação mestra, em que não há informação via uma dinâmica hamiltoniana, mas há informações probabilísticas sobre as transições entre estados do sistema. É evidente

que a validade de tal tratamento necessita ser mais bem situada, haja vista que a equação mestra descreve processos markovianos¹ o que nem sempre é o caso quando se estudam sistemas biológicos.

Inocentemente, assume-se que os sistemas biológicos representados pelas redes estão sujeitos a uma evolução temporal e que essa evolução temporal pode ser capturada parcialmente através de equações envolvendo probabilidades. É com essas equações simples que descreveremos os comportamentos que nomearemos não rigorosamente de equilíbrio e fora do equilíbrio.

É evidente que os sistemas vivos são sistemas abertos e que estão sempre em mudança constante. Essa noção sugere-nos que a descrição física desenvolver-se-ia no terreno da mecânica estatística fora do equilíbrio. Mas esse terreno é muito menos explorado do que o terreno de mecânica estatística no equilíbrio. Em tal caso, a tentativa de formalizar o estudo está mais longe ainda de ser alcançada e, por isso, a grande maioria dos trabalhos realizam um tratamento descritivo das redes ou então retiram informação através de modelagens computacionais. Apesar da falta de um formalismo físico de caráter geral, vários princípios estão sendo observados nas redes biológicas, o que vem contribuindo para o entendimento da biologia em sua visão sistêmica, a qual descreveremos em capítulos posteriores.

Tanto no equilíbrio como fora do equilíbrio, é interessante estudar o que acontece com as redes no caso em que algum regime estacionário seja alcançado. No caso de uma descrição via equação mestra

$$\frac{\partial P(y, t)}{\partial t} = \sum_{y'} P(y', t) w(y' \rightarrow y) - P(y, t) w(y \rightarrow y'),$$

em que $P(y, t)$ é a probabilidade de encontrar um sistema no estado y e $w(y' \rightarrow y)$ é interpretado como a probabilidade, na unidade de tempo, de que o sistema mude do estado y' para o estado y [22]. O regime estacionário é descrito impondo-se $\frac{\partial P(y, t)}{\partial t} = 0 = \sum_{y'} P(y', t) w(y' \rightarrow y) - P(y, t) w(y \rightarrow y')$. Usualmente utiliza-se da suposição extra do balanço detalhado em que cada termo do somatório se anula separadamente [22], ou seja,

$$P(y') w(y' \rightarrow y) = P(y) w(y \rightarrow y'). \quad (2.4)$$

Note-se que então a razão das taxas $w(y' \rightarrow y)/w(y \rightarrow y')$ fica relacionada à razão das distribuições estacionárias $P(y)/P(y')$.

De acordo com as considerações anteriores, discutiremos, nos próximos dois capítulos, alguns aspectos da teoria de redes no equilíbrio e fora do equilíbrio.

2.3 Redes sem-escala

2.3.1 A propriedade “sem-escala”

O tema dessa seção talvez merecesse um capítulo inteiro devido à importância que exerceu na orientação das pesquisas em redes complexas [15, 14, 4, 23, 24, 25, 26, 27]. Basicamente uma rede sem-escala é identificada pela sua distribuição de conectividades, que é da forma

$$P(k) \sim k^{-\gamma}. \quad (2.5)$$

¹Processos markovianos são processos sem memória, isto é, processos em que a probabilidade num tempo não depende das probabilidades anteriores [22]

Note que a conectividade média depende do expoente γ ,

$$\bar{k} = \sum_{k=1}^{\infty} k k^{-\gamma} \propto \int_1^{\infty} k^{-\gamma+1} dk = \frac{k^{-\gamma+2}}{-\gamma+2} \Big|_1^{\infty}.$$

Se a restrição de que a conectividade média seja finita for imposta, então deve-se ter $\gamma > 2$, que é a condição para que a série $\bar{k} = \sum_{k=1}^{\infty} k k^{-\gamma}$ convirja.

O termo “sem-escala” vem do fato de essa distribuição não apresentar um valor característico de k , isto é, a distribuição não possui uma escala característica. Isso pode ser evidenciado ao se comparar essa distribuição com uma do tipo Poisson, em que a maioria dos nós possuem conectividade em torno do valor médio. Como em uma rede sem-escala nós com conectividade alta estão presentes juntamente com nós com conectividade baixa, não há um valor característico de k , enquanto em uma distribuição do tipo Poisson é improvável que haja nós com conectividades diferentes do valor médio. No caso de Poisson o valor médio é o valor característico e no caso sem-escala não há valor característico.

As redes representam sistemas reais, que são sistemas finitos. Nem sempre a quantidade de nós é suficiente para que os nós de conectividade alta sejam estatisticamente incorporados na distribuição de conectividade. Isso impõe um corte nos valores de k acima do qual a distribuição já não mais reflete fielmente a amostra, caso em que a amostra possui pouquíssimos nós com $k > k_{cut}$. Dessa forma, o valor k_{cut} pode ser obtido igualando o número de nós com conectividade maior que k_{cut} a 1, i.e., $N \int_{k_{cut}}^{\infty} dk P(k) \approx 1$. Para $P(k) \propto k^{-\gamma}$, com $\gamma > 2$, assumindo que essa distribuição é válida para $k > k_0$, tem-se

$$k_{cut} \approx k_0 N^{1/(\gamma-1)}. \quad (2.6)$$

É evidente que, se N é suficientemente grande, o valor de corte garante que a distribuição é válida para valores de k grandes o suficiente para abarcar os *hubs*.

2.3.2 Outras redes sem-escala

Em nosso trabalho, estudaremos redes que representam sistemas biológicos, as quais apresentaremos no capítulo 3. Mas o objeto rede é usado para estudar vários outros sistemas, os quais apresentam a propriedade sem-escala.

A WWW (“World Wide Web”) é notadamente uma rede. Os nós são os documentos (páginas) e as ligações são os hiperlinks através dos quais se navega de uma página para outra. Como essa é uma rede direcionada, as distribuições diferenciam-se em $P_{in}(k) \propto k^{\gamma_{in}}$ e $P_{out}(k) \propto k^{\gamma_{out}}$. Em outubro de 1999, a Web consistia-se de 271×10^6 nós e 2130×10^6 hiperlinks. A conectividade média era $k_{in} = k_{out} = 7.85$. As distribuições de conectividade foram calculadas e realmente eram do tipo lei de potência, com expoentes $\gamma_{in} = 2.1$ e $\gamma_{out} = 2.7$ [28].

Redes sociais, em que os nós são pessoas e as ligações são formas de relacionamento, são também um excelente objeto para se aplicar à teoria de redes. Em particular, as redes em que as ligações representam colaboração constituem um campo interessante de estudos. As redes de colaboração podem representar as redes de amizade, redes terroristas, redes de atores de cinema etc.. Em particular, vejamos as redes de co-autorias de trabalhos científicos, em que os nós são autores de artigos e as ligações não direcionadas entre dois indicam se ambos os autores já foram co-autores de algum trabalho em comum [29]. Dois bancos de dados foram analisados: o Medline e o SPIRES (“Stanford Public Information Retrieval System”).

	N	\bar{k}	C	\bar{l}	l_{max}
Medline	1 520 251	18,1	0,006	4,6	24
SPIRES	56 627	173	0,73	4,0	19

Tabela 2.1: Algumas grandezas características das redes de co-autorias da Medline e SPIRES. N é a quantidade de nós. \bar{k} é a conectividade média. C é o coeficiente de aglomeração médio. \bar{l} é a distância média entre os nós e l_{max} é a maior distância entre dois nós.

Rede	N	L	γ	C	C/C_r	\bar{l}	\bar{l}/\bar{l}_r	Referências
WWW escanizadas pelo Altavista em outubro de 1999	$2,7 \times 10^8$	$2,1 \times 10^9$	$\gamma_{in}=2,1$ $\gamma_{out}=2,7$	-	-	16	-	[30, 28]
Home pages de uma companhia pública	4923	$1,3 \times 10^7$	$\gamma_{in}=2,05$	-	-	-	-	[31]
Citações do banco de dados ISI (1981-1997)	783 339	6 716 198	$\gamma_{in}=3,0$	-	-	-	-	[32]
Colaborações de atores de filme	212 250	61 085 555	2,3	-	-	4,54	1,25	[15]
Colaborações de autores de jornais de matemática	70 975	$0,132 \times 10^6$	2,1	0,59	$1,1^4$	9,5	1,16	[33]
Colaborações de autores de jornais de neurociência	209 293	$1,214 \times 10^6$	2,4	0,76	$1,4 \times 10^4$	6	1,2	[33]
Contatos sexuais	2 810	-	3,4	-	-	-	-	[34]
Cadeia alimentar em um parque	154	366	1	0,15	5	3,4	1,05	[35]
Chamadas de telefone	47×10^6	8×10^7	$\gamma_{in} = 2,1$	-	-	-	-	[36]
Rede de emails	4196	87 219	$\gamma_{in} = 1,5$	0,156	$3,25 \times 10^3$	4,95	0,48	[37]

Tabela 2.2: Medidas realizadas em algumas redes reais. N representa a quantidade de nós e L a quantidade de ligações. γ é o coeficiente da lei de potência que representa a distribuição de conectividade. C/C_r (\bar{l}/\bar{l}_r) representa a razão entre o coeficiente de aglomeração médio (distância média) da rede real e o coeficiente de aglomeração (distância média) de uma rede aleatória clássica com a mesma quantidade de nós e ligações (vide capítulo 4 para a definição da rede aleatória clássica).

Em ambas as redes, há uma grande componente conexa (componente em que todos os nós se encontram conectados entre si), à qual pertencem cerca de 90% dos nós. Algumas medidas foram também obtidas e estão na tabela 2.1. As distribuições de conectividade da Medline não podem ser escritas em termos de uma única lei de potência. Já a distribuição da SPIRES é do tipo lei de potência com $\gamma = 1, 2$. Esse é um exemplo que mostra que as redes reais apresentam efeitos de tamanho finito. Se $\gamma < 2$, deveríamos ter $\bar{k} = \infty$, o que não é verdade, porque a rede é finita. Apesar de haver discussões em torno da validade da lei de potência como sendo a relação funcional que representa a distribuição, o fato é que essas distribuições possuem decaimento lento, o que possibilita a existência de nós altamente conectados, os hubs.

A título de ilustração, vejamos alguns outros exemplos de redes que já foram estudados. As redes de telefonemas, em que os nós são números de telefones e as ligações direcionadas indicam que um número realizou uma chamada em direção a outro número. As redes de e-mail podem ser construídas de maneira similar, as redes de co-participação em filmes, que são mais um exemplo de redes de colaboração. Em particular, as redes biológicas oferecem um amplo campo de aplicação da teoria de redes. Como as redes biológicas são o principal tema do presente trabalho, abstenho-nos de maiores comentários, os quais deixamos para o capítulo 3. Algumas medidas, obtidas a partir do estudo de algumas redes reais, podem ser vistas na figura 2.2.

A propriedade sem-escala parece ser uma propriedade ubíqua nas redes reais. A tabela 2.2 mostra alguns valores característicos obtidos a partir de redes reais.

2.3.3 Propriedades notáveis

A característica mais notável de uma rede sem-escala é a presença de nós que possuem muito mais conexões do que a média. Tais nós recebem o nome de *hubs*. A presença dos *hubs* confere propriedades interessantes

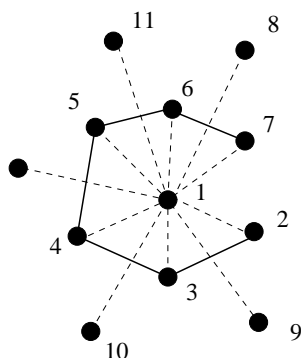


Figura 2.3: O nó 1 é um hub da rede. Na ausência do hub, i.e., na ausência das ligações tracejadas, a distância entre os nós 2 e 7 seria 5. O hub cria um atalho reduzindo a distância para 2.

às redes. Vejamos algumas propriedades.

Redes sem-escala são conhecidas como "mundos ultra pequenos". A distância entre pares de vértices é drasticamente diminuída, se houver a presença de hubs, pois eles fornecem atalhos na rede. Em outras palavras, como um mesmo hub se liga a vários vizinhos, esses vizinhos passam a ficar, no máximo, a uma distância 2 uns dos outros. A figura 2.3 ilustra essa propriedade. Um estudo um pouco mais detalhado será discutido na seção 4.3.

Outra propriedade é conhecida como tolerância ao erro [38]. Para avaliar essa medida, verifica-se como as distâncias médias entre os nós variam em função da remoção aleatória de uma percentagem f de nós. A distância média quantifica de certa forma a habilidade de dois nós se comunicarem e a remoção dos nós pode ser entendida como o mecanismo que introduz erro na comunicação. A tolerância da rede sem-escala é comparada com a de uma que possui distribuição de conectividade tipo Poisson. Constata-se que, nessa última, a distância média cresce linearmente com f enquanto na rede sem-escala a distância média permanece praticamente inalterada no intervalo $0\% \leq f \leq 5\%$. Isso mostra que a rede sem-escala é tolerante em relação à remoção aleatória de nós. Em contrapartida, se a remoção for predeterminada de tal forma que os hubs sejam os alvos, então as redes sem-escala são extremamente vulneráveis.

Uma propriedade muito mais sutil é a capacidade que as redes sem-escala têm de alterar as suas ligações a fim de atingir um objetivo preestabelecido [16]. De uma maneira que só ficará claro na seção 6.1, pode-se associar à rede um sistema dinâmico, cujo comportamento depende da topologia da rede. Fixa-se, como objetivo, um determinado comportamento da dinâmica e então se pede que, por meio de redirecionamentos aleatórios das ligações, alcance-se o objetivo. É evidente que, durante o processo de redirecionamento, a distribuição de conectividade não pode ser alterada. Constata-se que as redes sem-escala aproximam-se do objetivo muito mais rapidamente que uma rede com distribuição tipo Poisson. Isso mostra que as redes sem-escala têm maior capacidade de evolução rumo a objetivos predeterminados.

As redes sem-escala estão bastante presentes na natureza, como veremos no capítulo 3 no caso das redes biológicas. A escolha de tal topologia talvez esteja refletida no ganho evolucionário que suas propriedades fornecem às espécies. É evidente que uma topologia que possua maior capacidade de evoluir rumo a um novo objetivo, tem muito mais chance de se manter presente no cenário da evolução em vista do processo de seleção natural. A presença de distribuição de conectividade do tipo lei de potência define, portanto, uma importante classe de redes: as redes sem-escala.

Capítulo 3

Redes biológicas

É indubitável que estamos cercados de coisas que chamamos de “coisas vivas”, as quais apresentam um ciclo de vida durante o qual atividades particulares garantem sua existência. Mas também é totalmente consensual que as unidades que compõem uma célula e executam suas atividades são “coisas mortas”. Um composto de carbono, um lípido, por exemplo, é unanimemente considerada uma “coisa morta”, ao passo que, uma estrutura complexa formada por várias estruturas de carbono, um organismo humano, por exemplo, manifesta um comportamento que classificamos como vivo. Uma concepção que é defendida por alguns pesquisadores da origem da vida é: nada na célula é vivo exceto a célula inteira. Essa perspectiva sugere a necessidade de uma abordagem que capture aspectos globais e como as partes se interagem a fim de executar uma tarefa determinada. Em outras palavras, é tão importante estudar as propriedades de uma determinada proteína quanto conhecer como as várias proteínas interagem entre si a fim de que haja a coordenação, por exemplo, dos processos metabólicos.

Um sistema vivo é composto por uma quantidade grande de unidades “mortas” que interagem entre si de acordo com as leis físicas compondo uma maquinaria celular complexa. Essas unidades estão em entrada e saída constante do organismo vivo: a composição de um organismo humano, por exemplo, está em renovação constante, em que matéria-prima nova adquirida pela alimentação é utilizada na substituição de células que vão se decompondo. O que se mantém praticamente constante, durante o percurso de um ser vivo, são as informações sobre o funcionamento dele próprio, que se encontram armazenadas em um biblioteca imensa conhecida como genoma. Essa biblioteca é formada por livros conhecidos como DNA, que contém as receitas para executar todas as atividades que compõem o ser vivo.

A fim de que as informações genéticas sejam executadas na forma de processos biológicos, é necessário uma maquinaria complexa. Essa maquinaria complexa é realizada pelas interações entre as várias unidades que compõem o sistema vivo, tais como DNA, RNA, as proteínas e as pequenas moléculas, em conjugação com as informações provindas do meio externo. Essas interações, como era de se esperar, podem ser representadas na forma de um grafo, compondo as redes celulares. São as redes celulares, portanto, que representam a maquinaria complexa que executa as instruções contidas nos genoma transformando-as em atividades biológicas através da qual a vida se expressa, i.e., não são as unidades celulares isoladas que executam as instruções, mas todas as unidades que, funcionando em conjunto, executam-nas.

As redes presentes em uma célula ligam-se a redes em outras células formando possivelmente uma estrutura hierárquica, como, por exemplo, a hierarquia entre o sistema nervoso central e as células em vários órgãos que estão submetidas aos comandos provenientes do sistema nervoso central. É evidente que,

apesar de um organismo ser uma unidade bem definida, ele se encontra inserido em um contexto biológico maior, em que as interações são conhecidas como interações ecológicas. Um exemplo de tal interação é a relação predador-presa, cuja representação em termos de redes é conhecida como cadeia alimentar [39]. Gostaríamos de deixar claro que o termo redes biológicas não se restringe às redes celulares, mas também a níveis mais amplos de relações entre os organismos. Contudo neste trabalho somente estudaremos as redes intracelulares. Nas próximas seções apresentamos alguns exemplos de redes celulares [1] no que diz respeito a seus aspectos estruturais ou topológicos. A maneira pela qual as redes executam os processos biológicos será tratada no capítulo 6.

3.1 Rede de interação entre proteínas

As proteínas são os principais elementos organizadores das funções celulares. Elas agem como enzimas, elementos estruturais, mensageiras de sinais e máquinas biológicas. Muitas vezes, para realizar suas funções, as proteínas se ligam umas às outras formando complexos protéicos. Em vista disso, pode-se definir a rede de proteínas. Nas redes de interação entre proteínas, os nós são as proteínas e dois nós estão conectados por uma ligação não direcionada se as duas proteínas correspondentes se ligam fisicamente.

As técnicas utilizadas na construção das redes nem sempre utilizam procedimentos baseados propriamente em testes físicos de interação, isto é, nem sempre se colocam duas proteínas em laboratório a fim de observar se há ou não interação. Comentemos alguns métodos de construção das ligações baseados em inferências [8]. Inicialmente obtém-se a função de algumas proteínas via experimentos bioquímicos, genéticos ou estruturais. Mas, a partir do conhecimento da função de uma proteína busca-se, através de algumas inferências biologicamente razoáveis, descobrir proteínas funcionalmente relacionadas às que já têm uma função bem determinada. A conjectura extra é a de que proteínas funcionalmente relacionadas costumam interagir entre si. Com isso atribuem-se ligações entre as proteínas cuja relação foi inferida. Vejamos algumas dessas inferências. Procura-se proteínas que possuam seqüências de aminoácidos semelhantes à proteína original e atribui-se uma ligação entre essas proteínas. Esse é conhecido como método de homologia. Outro método é o perfil filogenético, em que se levanta o padrão de presença ou ausência de uma proteína em vários organismos de espécies distintas. Se duas proteínas têm o mesmo perfil de presença e ausência infere-se que possuem uma ligação entre si. Pode-se também pesquisar se duas proteínas em um organismo estão presentes como uma proteína fundida em algum outro organismo. Esse método é conhecido como seqüência Rosetta Stone. Ainda existem outros métodos, mas todos, embora com alguma justificativa biológica que os valide, ainda fornecem redes cujas as ligação refletem relações funcionais a partir das quais pode-se inferir possíveis interações físicas.

Duas proteínas estão funcionalmente ligadas se elas interagem diretamente ou se participam seqüencialmente de um mesmo caminho biológico. Essa última espécie de ligação pode ser melhor evidenciada através de um exemplo. Consideram-se as reações químicas $A \rightarrow B$ e $B \rightarrow C$ mediadas pelas enzimas α e β respectivamente. Em vista desse caminho metabólico, atribui-se uma ligação funcional $\alpha \rightarrow \beta$. Algumas vezes essa ligação corresponde a uma interação física, como no caso de complexos multienzimáticos (várias enzimas ligadas fisicamente formando um complexo).

Mas, de qualquer forma, é importante ter claro que tipo de interação se está considerando na construção das redes, pois em alguns casos a ligação representa uma interação física [40, 41] e em outros casos representa ligações funcionais obtidas através de inferências [8].

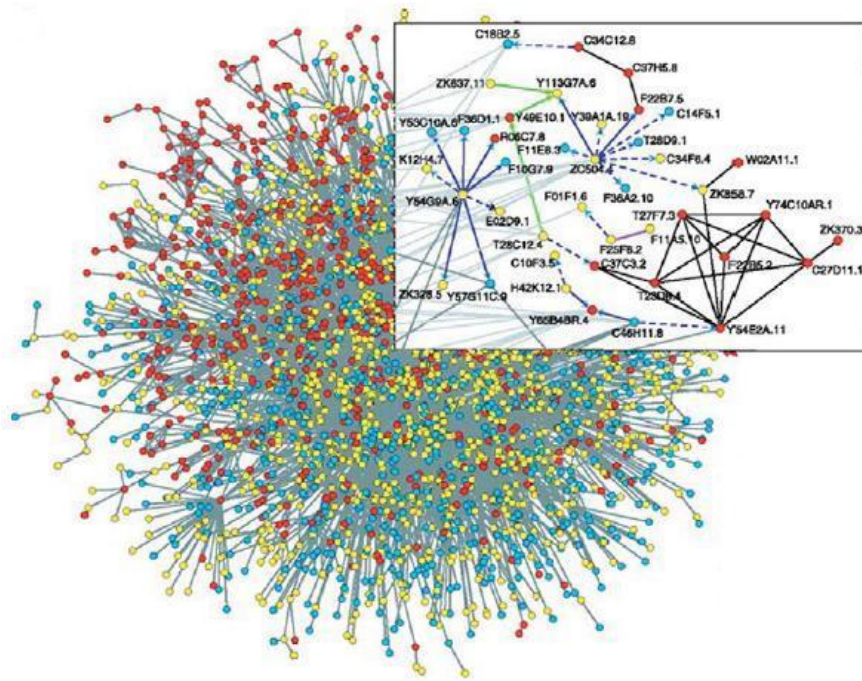


Figura 3.1: Rede de interação de proteínas da *C. elegans*. A caixa em destaque apresenta uma parte da rede em detalhes. As interações foram obtidas a partir de testes de interação física e a partir de inferências similares às descritas na seção 3.1. [1]

Devido aos métodos que atribuem ligações entre proteínas por inferências, a rede poder apresentar ligações que *in vivo* não representariam uma ligação. Essas ligações são chamados falsos positivos. Apesar da existência de falsos positivos, a análise da estrutura dessas redes revela a existência de uma grande componente conexa (veja figura 3.1, cujos nós encontram-se não muito distantes uns dos outros (como veremos no próximo capítulo, essa rede é um ‘mundo pequeno’). A distribuição de conectividades é aproximadamente do tipo lei de potência e a distribuição do coeficiente de aglomeração em função da conectividade é aproximadamente $C(k) = B/k^\beta$ (Veja a figura 3.2). Essa lei de potência indica que a vizinhança de um nó muito conectado é um tanto quanto esparsa.

3.2 Redes metabólicas

Desde que os sistemas vivos confeccionaram um membrana que os tornou um sistema individualizado, foi necessário que ele próprio construísse a sua maquinaria de transformação de energia e elaboração de materiais para se manter estruturalmente. Em vista dessa necessidade, uma máquina complexa foi montada: os processos metabólicos. Basicamente, tais processos envolvem reações químicas entre metabólitos que são catalisados por enzimas. Pode-se representar esses processos em uma rede de várias formas. Uma forma simples [2] desta representação pode ser explicada através do seguinte exemplo. Considera-se a reação $A + B \rightleftharpoons C$. Os nós representarão os metabólito A, B e C . As ligações serão $A - C$ e $B - C$. Dessa

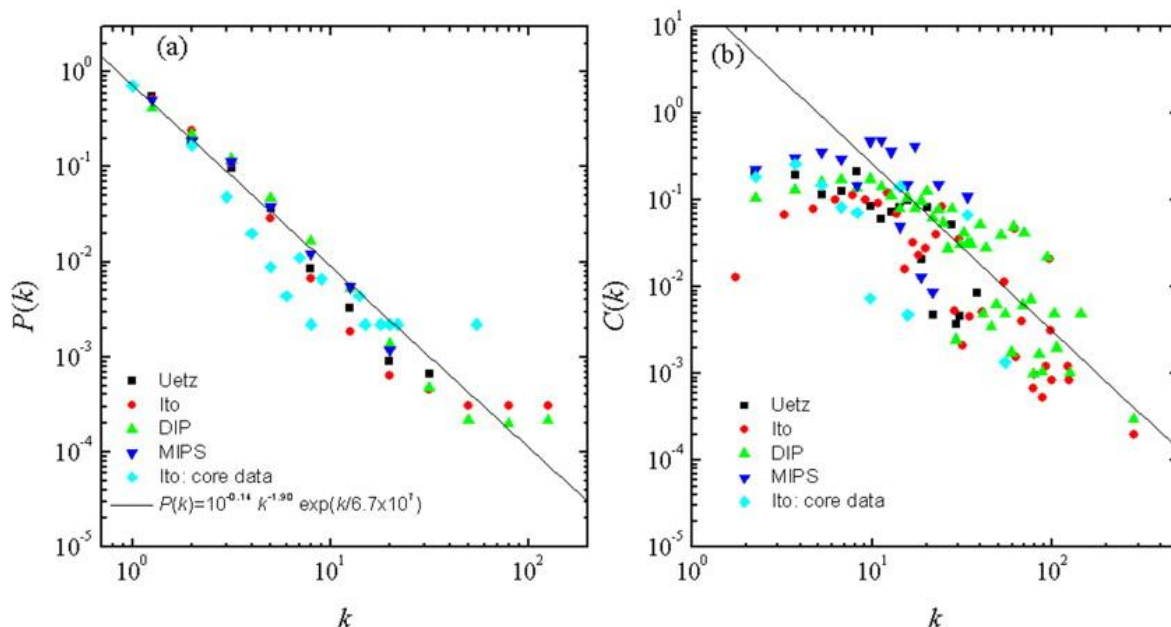


Figura 3.2: Propriedades topológicas da rede de interação de proteínas da *S. Cerevisiae* obtidas a partir de quatro bancos de dados. (a) Distribuição de conectividade. A linha sólida representa a distribuição $P(k) \propto k^{-2.5}$. (b) Coeficiente de aglomeração. A linha sólida representa a função $C(k) = \beta/k^2$ [1]

forma, a regra geral é: para cada metabólito atribui-se um nó e, se dois nós participam da mesma reação, um como reagente e o outro como produto, liguem-se ambos à rede. Outras representações podem ser feitas, em que as enzimas fazem parte da rede, ou mesmo em que as reações são consideradas como nós que possuem ligações entre si, se compartilham algum metabólito. Uma componente da rede metabólica da *E. coli* é mostrada na figura 3.3

Em geral, todas as representações da rede metabólica indicam que são aproximadamente sem-escala e os coeficientes de aglomeração distribuem-se segundo $C(k) \cong 1/k$. Essas representações indicam também uma distância média pequena entre os nós e que essa distância média é independente dos organismos. A partir das redes metabólicas pode-se fazer vários estudos estruturais envolvendo fluxos metabólicos [42], relações funcionais entre a estrutura e os caminhos metabólicos conhecidos [43, 11] etc..

3.3 Redes de transcrição genética

As informações contidas no genoma possibilitam a produção de proteínas que, como vimos, realizam papel preponderante na execução das funções biológicas. A fim de que um processo biológico ocorra, as ações dos genes devem ser reguladas para que a produção de RNA e proteínas ocorra de maneira a coordenar corretamente o dado processo. Em geral, os genes podem ser ativados ou inibidos através da atuação de fatores de transcrição ou inibição. Pode-se representar esse esquema de regulação através de uma rede em que os nós são os genes e as ligações representam ativação ou inibição do gene. O significado da ligação pode ser melhor entendido no seguinte exemplo. Seja $A \rightarrow B$ uma ligação de ativação. Isso significa que a proteína transcrita a partir gene A atua como fator de ativação do gene B. Um exemplo de rede de transcrição é mostrado na figura 3.4

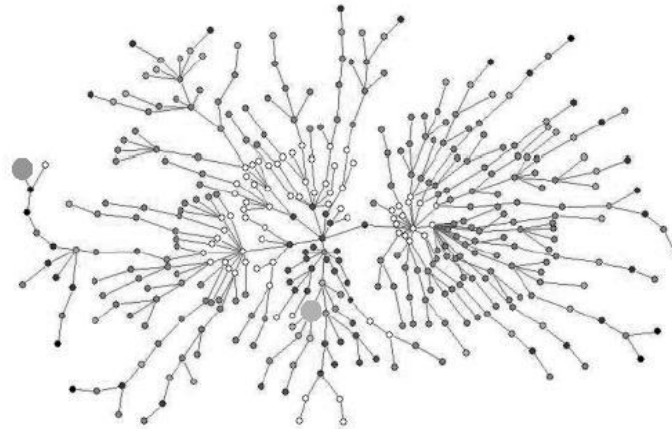


Figura 3.3: Componente conexa da rede metabólica da *E. Coli* que contém a glicose. O nó maior na região central representa a glicose e o nó maior na região mais à esquerda representa o nó mais distante da glicose. [2]

Essas redes regulatórias serão objeto de um estudo mais detalhado nos capítulos finais. Por isso nos abstermos de maiores comentários limitando-nos apenas a dizer que tais redes em geral são sem-escala [44].

3.4 Redes de tradução de sinais

Tanto no nível de uma célula quanto no de um organismo inteiro, as partes dos sistemas se integram com o meio em que estão inseridos. Por exemplo, os seres aeróbicos retiram oxigênio do meio e as células das glândulas no organismo humano recebem informação do sistema nervoso central a fim de iniciar a síntese de hormônios. Não importa o nível de organização que se está considerando, os sistemas vivos evoluíram integrados com o meio. Essa necessidade deu origem às redes de tradução de sinais, cujos mecanismos conectam sinais extracelulares aos sistemas de controle dos fatores de transcrição. Os mecanismos que executam a tradução dos sinais são usualmente cadeias de reações químicas e interação entre proteínas que se iniciam com moléculas que capturam o sinal extracelular. A representação em termos de redes dos processos de tradução de sinais envolve várias espécies de interações: reações químicas, interação proteína-proteína entre outras, sendo, por isso, uma miscigenação dos tipos anteriormente citados de redes.

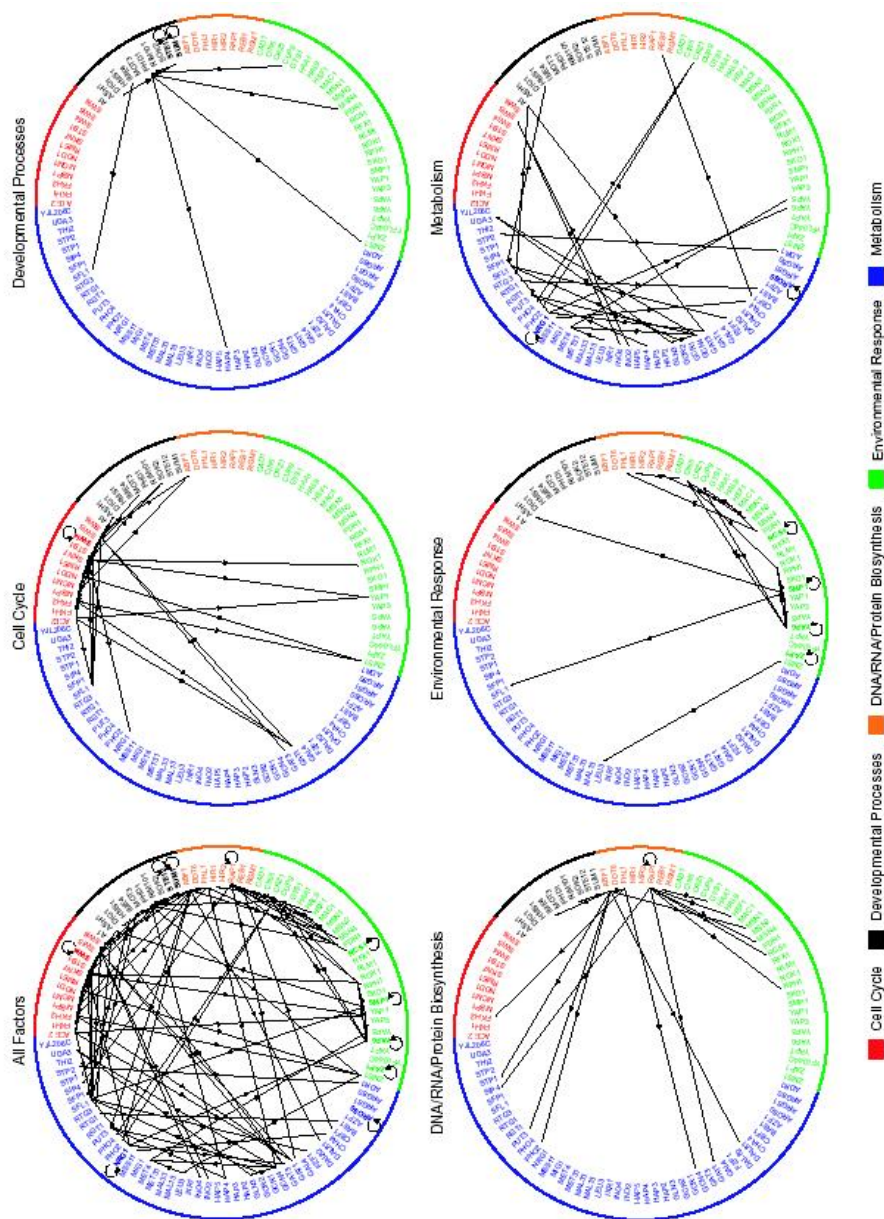


Figura 3.4: Rede de reguladores da transcrição da *S. Cerevisiae*. Os nós são os fatores de transcrição e uma ligação do tipo $A \rightarrow B$ significa que o fator A atua na transcrição do fator B . As seis figuras representam os mesmos nós, mas cada figura separadamente contém as ligações responsáveis pela coordenação de algum processo específico. A figura no canto inferior contém as interações de todos os processos. [3]

Capítulo 4

Redes no equilíbrio

Esse capítulo dedica-se ao estudo teórico de redes que Mendes et al. [20] classificam como redes no equilíbrio. Essa classificação não possui critérios formais, mas apela para uma noção intuitiva de equilíbrio. Considera-se um sistema físico em equilíbrio, se algumas propriedades macroscópicas são constantes e uniformes. A constância e uniformidade de tais propriedades não refletem, necessariamente, a realidade microscópica em que fluxos locais podem estar presentes em regime de equilíbrio. Nesse mesmo sentido, iremos descrever um processo dinâmico em que as ligações de uma rede são constantemente modificadas, mas que atingem um regime de equilíbrio quando certa caracterização macroscópica é alcançada e permanece constante e uniforme.

Sistemas, cujas configurações sejam descritas em termos probabilísticos, usualmente podem ser representados por ensembles, cuja contagem de elementos numa dada configuração fornece a probabilidade de se obter a configuração. Tais ensembles em geral são construídos através de algum processo dinâmico que possua um regime de equilíbrio no limite de tempos longos ($t \rightarrow \infty$). Seguindo esse tratamento típico da mecânica estatística de equilíbrio, define-se um ensemble estatístico de redes como

1. um conjunto G de grafos,
2. uma regra que associa algum peso estatístico $P(g) > 0$ para cada grafo g do conjunto G .

A regra que associa o peso estatístico pode ser obtida de várias maneiras:

1. Pode-se atribuir um peso estatístico diretamente a cada grafo $g \in G$.
2. Pode-se definir um processo no conjunto G que resulte em pesos estatístico para os grafos.

Em geral tal construção nos possibilitaria inferir valores médios de quantidades que dependem da configuração das redes. Seja $S(g)$ tal quantidade, então

$$\langle S \rangle = \frac{1}{Z} \sum_{g \in G} S(g)P(g),$$

em que $Z \equiv \sum_{g \in G} P(g)$. Contudo nos concentraremos em estudar principalmente a distribuição de conectividade

$$P(k) = \langle N(k) \rangle / N, \tag{4.1}$$

em que $\langle N(k) \rangle$ é o número médio de vértices com conectividade k . O motivo de um maior enfoque nessa distribuição vem da importância em se classificarem as redes dentro da classe de redes sem-escala.

Note-se que a atribuição pré-determinada de um peso estatístico pode ser útil, quando se quer construir uma rede que obedeça a certa propriedade, como, por exemplo, certa distribuição de conectividade, mas que seja totalmente aleatória nos outros aspectos. Na análise de sistemas biológicos, tais construções são de extrema importância na verificação da relevância estatística de correlações na rede. Contudo o estudo de regras dinâmicas que conduzem a uma determinada configuração das redes possui uma motivação extra, a possibilidade de se estudarem processos evolucionários reais como fatores de formação da estruturas gerais das redes complexas.

A título de clareza de notação, a notação $\langle X \rangle$ refere-se a médias de uma quantidade X no ensemble. A única média no ensemble que utilizaremos é a quantidade média $P(k)$, de acordo com a equação 4. Já os valores médios obtidos a partir da distribuição $P(k)$ serão representados através de uma barra, como em \bar{X} , em que $\bar{X} = \sum_k X(k)P(k)$. Começemos inicialmente pela análise de um ensemble de redes aleatórias clássicas.

4.1 Redes aleatórias clássicas

Embora a teoria de grafos seja bem anterior ao século XX, foram os matemáticos Erdos e Renyi que introduziram o conceito de grafos aleatórios. Basicamente tais grafos são construídos a partir de um processo aleatório em que dois vértices do grafo estarão ligados com probabilidade p e não ligados com probabilidade $1 - p$. Essa teoria é bem estudada e resultados rigorosos podem ser encontrados em [19]. Apenas nos limitaremos a determinar a distribuição de conectividade de um rede aleatória clássica com N vértices.

Um dado vértice estará ligado a k outros vértices com probabilidade p^k e estará não ligado aos $N - 1 - k$ vértices restantes com probabilidade p^{N-1-k} . Como os k possíveis vizinhos podem ser escolhidos entre $N - 1$ vértices, a probabilidade de um dado vértice estar ligado a k outros vértices é

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

Vamos mostrar que a conectividade média é $\bar{k} = p(N-1)$ e que o grafo possui em média $pN(N-1)/2$ ligações. Pela definição temos

$$\bar{k} = \sum_{k=0}^{N-1} k \binom{N-1}{k} p^k q^{N-1-k} = p \frac{\partial}{\partial p} \sum_{k=0}^{N-1} \binom{N-1}{k} p^k q^{N-1-k},$$

em que $q = 1 - p$. Por sua vez o somatório vale exatamente $(p+q)^{N-1}$, que pela definição de q vale 1. Com isso,

$$\bar{k} = p \frac{\partial}{\partial p} (p+q)^{N-1} = (N-1)p(p+q)^{N-2} = p(N-1).$$

Se somarmos a conectividade de todos os nós, obteremos o dobro da quantidade de ligações. A quantidade

média de ligações \bar{L} é portanto

$$\bar{L} = \frac{\sum_{k=1}^{N-1} NkP(k)}{2} = \frac{N\bar{k}}{2} = \frac{pN(N-1)}{2}.$$

É também conhecido que, no limite $N \rightarrow \infty$ para \bar{k} fixo, a distribuição de conectividade assume a distribuição de Poisson

$$P(k) = e^{-\bar{k}} \bar{k}^k / k!.$$

É possível calcular facilmente algumas grandezas, como o coeficiente médio de aglomeração e a distância média entre pares de nós. O coeficiente médio de aglomeração é $\bar{C} = 2y/z(z-1)$, em que y é a quantidade média de ligações entre os vizinhos do nó i e z é a quantidade média de vizinhos. Mas em uma rede aleatória com N nós e L ligações, a probabilidade de que dois nós estejam conectados é $2L/N^2$. Como $z = 2L/N$, temos $2L/N^2 = z/N$ e, com isso, a quantidade média de ligações entre os vizinhos é

$$y = \frac{z(z-1)}{2} \left(\frac{z}{N} \right)$$

e, obviamente, $z = \bar{k}$. Substituindo essas expressões obtém-se

$$\bar{C} = \bar{k}/N.$$

Existe uma maneira fácil de fazer o cálculo da distância média, desde que seja assumido que a rede é do tipo árvore. Uma árvore é uma rede conexa em que não há nenhum caminho fechado [19]. Nesse caso, a quantidade média de vizinhos de cada nó é \bar{k} e a quantidade de nós a uma distância igual ou menor que l de um dado nó é \bar{k}^l , como pode ser visto na figura 4.1. Um estimativa para a quantidade de nós a uma distância arbitrário de um dado nó é $\bar{k}^{\bar{l}}$, em que \bar{l} é a distância média entre pares de nós. Mas essa quantidade é justamente a quantidade total N de nós, i.e., $N \approx \bar{k}^{\bar{l}}$. Invertendo essa equação obtém-se

$$\bar{l} \approx \ln N / \ln \bar{k}.$$

4.2 Construção de redes em equilíbrio

Como dissemos na seção anterior, pode-se tanto atribuir um peso estatístico predeterminado, como impor uma regra de evolução para as redes em um ensemble de maneira tal que certa distribuição de conectividades seja alcançada. A segunda forma é mais interessante, pois na realidade sabemos que, mesmo durante o equilíbrio, existem fluxos locais no espaço de fase de sistemas mecânicos que obedecem ao princípio de balanço detalhado e gostar-se-ia, da mesma forma, de impor um fluxo na rede que satisfaça as condições de equilíbrio.

No caso de atribuição de um peso estatístico pré-determinado a um ensemble, cuja distribuição de conectividade seja preestabelecida, pode-se construir um ensemble da seguinte forma:

- Seja $N(j)$ uma seqüência de inteiros não negativos tal que $\sum_j N(j) = N$. O ensemble será então formado por todos os grafos de tamanho N tal que a quantidade de vértices com conectividade k é

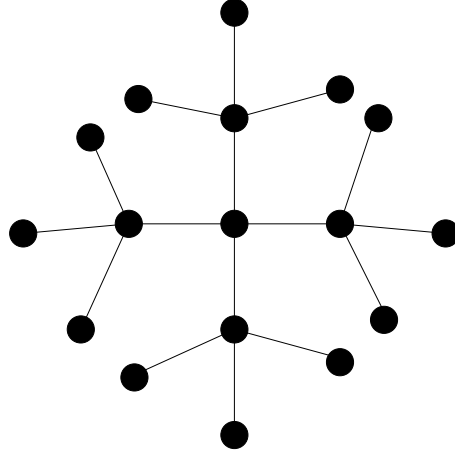


Figura 4.1: Ilustração de um grafo do tipo árvore. Note a ausência de caminhos fechados.

$N(k)$.

- Atribuem-se a todos os grafos pesos estatísticos iguais

No caso em que uma regra dinâmica é imposta, pode-se construir o ensemble da seguinte forma [20]:

- Considere-se o conjunto de grafos que possuem uma quantidade L fixa de ligações.
- A cada passo da evolução, o término de uma ligação escolhida aleatoriamente é redirecionada para um vértice com conectividade k com taxa $f(k)$. No limite $t \rightarrow \infty$ o ensemble estará definido através da seqüência $\{f(k)\}$ e L (ou, equivalentemente, $\bar{k} = 2L/N$).

Ambas as construções acima não são únicas, mas fornecem um meio de se obter a distribuição de conectividade através de equações mestras, como faremos logo a seguir. Note-se também que, na regra de evolução temporal, está presente a noção de ligação preferencial tanto no momento da escolha do vértice que receberá a ligação, com probabilidade relacionada a $\{f(k)\}$, quanto no momento da escolha do vértice que será selecionado para perder uma ligação com probabilidade $k_j/(2L)$. A noção de ligação preferencial é importante, pois o redirecionamento das ligações não é feito de maneira aleatória, mas dando preferência aos nós com conectividade k de acordo com a função preferência $f(k)$. Note-se também que, com isso, é possível conceber a possibilidade de um regime de equilíbrio. Da mesma forma que um nó com conectividade k perde uma ligação com probabilidade que depende de k , um outro nó ganhará essa ligação perdida com probabilidade que depende novamente de k . Se essas dependências em k se equilibrarem pode-se obter um regime de equilíbrio.

A equação mestra para a evolução temporal das distribuições $P(k, t) = \langle N(k) \rangle / N$, com $\sum_k \langle N(k, t) \rangle = N$, de acordo com a regra dada é

$$\langle N(k, t+1) \rangle = \langle N(k, t) \rangle - \frac{f(k)}{N \langle f(k) \rangle} \langle N(k, t) \rangle + \frac{f(k-1)}{N \langle f(k) \rangle} \langle N(k-1, t) \rangle - \frac{k}{N \bar{k}} \langle N(k, t) \rangle + \frac{k+1}{N \bar{k}} \langle N(k+1, t) \rangle, \quad (4.2)$$

em que $\langle f(k) \rangle = \sum_k P(k, t) f(k, t)$ (omitimos a dependência em t na expressão de $\langle f(k, t) \rangle$ pois estaremos interessados apenas no regime estacionário). Analisemos cada termo.

2º termo: a probabilidade de um vértice com conectividade k ser escolhido para receber uma nova ligação é $Af(k)$, em que A é uma constante. A cada passo somente um vértice é escolhido, de forma que $A \sum_k N(k) f(k) = 1$. Mas como $N(k) = NP(k)$, tem-se $A = 1/(N \langle f(k) \rangle)$. A taxa com que um vértice de conectividade k ganha uma nova ligação é $f(k)/(N \langle f(k) \rangle)$ e a quantidade de vértices que deixam de ter k ligações nesse processo é o segundo termo da expressão.

3º termo: um vértice que possua $k-1$ ligações ganhará uma nova ligação a uma taxa $f(k-1)/(N \langle f(k) \rangle)$. A quantidade de vértices que tinham $k-1$ ligações e passam a ter k ligações é portanto o terceiro termo.

4º termo: quando se escolhe aleatoriamente uma ligação para ser redirecionada, o vértice que perde a ligação o faz com probabilidade $k/(2L)$. Como $2L = \bar{k}N$, obtém-se a taxa $k/(N\bar{k}) \langle N(k, t) \rangle$ com que um vértice com k ligações passa a ter $k-1$.

5º termo: quando se escolhe um vértice para perder uma ligação, se esse vértice possuir $k+1$ ligações, então ele irá contribuir para o aumento de vértices com k ligações. Esse aumento está representado no quinto termo.

Substituindo $\langle N(k) \rangle$ por $P(k)$, obtém-se no limite contínuo

$$N \frac{\partial P(k, t)}{\partial t} = \frac{[-f(k)P(k, t) + f(k-1)P(k-1, t)]}{\langle f(k) \rangle} + \frac{[-kP(k, t) + (k+1)P(k+1, t)]}{\bar{k}}.$$

Reordenando os termos dessa equação obtém-se

$$N \frac{\partial P(k, t)}{\partial t} = \left[\frac{1}{\langle f(k) \rangle} f(k-1)P(k-1, t) - \frac{kP(k, t)}{\bar{k}} \right] + \left[\frac{k+1}{\bar{k}} P(k+1, t) - \frac{f(k)}{\langle f(k) \rangle} P(k, t) \right].$$

A condição do balanço detalhado 2.4 fornece que:

$$\frac{f(k)}{\langle f(k) \rangle} P(k) - \frac{k+1}{\bar{k}} P(k+1) = \text{const} = 0. \quad (4.3)$$

A relação acima sugere duas análises distintas. Dada uma distribuição $P(k)$ de equilíbrio, determina-se quais funções $f(k)$ são capazes de gerar essa distribuição, ou então, dada uma função $f(k)$ determina-se a distribuição de equilíbrio $P(k)$. Por exemplo, se $P(k) \propto \exp(-k)$, então, de acordo com a equação 4.3, tem-se $f(k) = e^{-1}(k+1)\langle f(k) \rangle/\bar{k}$, isto é, $f(k) \propto k+1$. Se $P(k) \propto k^{-\gamma}$ então pela equação 4.3 deve-se ter

$$f(k) = \frac{(k+1)(k+1)^{-\gamma} \langle f(k) \rangle}{k^{-\gamma} \bar{k}}.$$

No limite $k \rightarrow \infty$ tem-se

$$\frac{\bar{k}f(k)}{\langle f(k) \rangle} \sim k + o(k) \quad (4.4)$$

($o(k)$ significa potências de k menores que 1). Essa última relação é uma condição necessária para que a distribuição estacionária seja do tipo lei de potência.

Suponha $f(k) \sim k + o(k)$ no limite $k \rightarrow \infty$, i.e., $f(k)$ é assintoticamente linear em relação a k . Para que $P(k)$ seja uma lei de potência é necessário que a equação 4.2 seja válida. Mas essa equação só é idêntica a $f(k) \sim k + o(k)$ para o único valor crítico

$$\bar{k}_c = \langle f(k) \rangle. \quad (4.5)$$

Em outras palavras, para se obter uma distribuição $P(k)$ do tipo lei de potência são necessárias duas condições: (i) a função preferência deve assintoticamente linear em k e (ii) $\bar{k}_c = \langle f(k) \rangle$.

Qual é então o significado de um ensemble em que $\bar{k} \neq k_c$? Dedicemos uma subseção para elucidar essa questão.

4.2.1 Condensação

A relação 4.5 parece indicar que se $f(k) = k$ então sempre teremos a condição necessária para que ocorra um distribuição tipo lei de potência, pois nesse caso $k_c = \bar{k}$. Entretanto suponha-se a seguinte situação: suponha $P(k) \sim k^\gamma$ a priori, sem levar em consideração a quantidade L de ligações disponíveis na rede com N nós, e também fixe $f(k) = k$. Em relação a $P(k)$, já com a constante de normalização $\gamma - 1$ o valor médio da conectividade é

$$\bar{k} = \int_1^\infty k \frac{k^{-\gamma}}{\gamma - 1} dk = \frac{1}{(\gamma - 1)(\gamma - 2)}. \quad (4.6)$$

Para N e L fixos, é sempre verdade que $\bar{k} = 2L/N$. Mas para que a equação 4.2.1 seja consistente devemos ter

$$\bar{k} = \frac{2L}{N} = \frac{1}{(\gamma - 1)(\gamma - 2)} \quad (4.7)$$

Isso mostra que uma dada distribuição de conectividade só é consistente para um valor de \bar{k} .

No processo de construção das redes em equilíbrio, a distribuição das ligações é obtida através de um processo de redirecionamento, que é determinado pela função preferência $f(k)$. Suponha-se que a rede possua 100 nós, mas apenas 1 ligação. Nesse caso temos que $\bar{k} = 1/50$ e é evidente que $P(k) = \delta_{k,1}$. Vejamos, agora, o caso geral construído a partir de uma função preferência $f(k)$ em que a distribuição estacionária de conectividade é do tipo $P(k) \propto k^{-\gamma}$. Como $f(k)$ é fixo, a distribuição de ligações só alcançará o valor estacionário se houver ligações suficientes para que os nós recebam-nas e organizem-se de tal forma a exibir a distribuição desejada. Além disso, é necessário que a quantidade de ligações não ultrapasse um determinado valor, pois nesse caso teríamos nós com uma quantidade muito maior de ligações do que a que seria encontrada, caso a distribuição desejada estivesse presente.

Os argumentos apresentados na seção anterior mostram que se $P(k)$ é do tipo lei de potência, então é necessário que $f(k)$ seja assintoticamente linear e que $\bar{k} = \langle f(k) \rangle$. Essa última relação fornece o valor crítico para o qual a distribuição $P(k)$ é consistente com a função preferência e com a quantidade de ligações e nós. Mas, então, o que ocorre se a mesma função preferência for utilizada e $\bar{k} \neq k_c$?

A análise dos regimes $k < k_c$ e $k > k_c$ pode ser formalizada através de um problema análogo, conhecido como modelo “Backgammon” [20, 45]. Esse modelo basicamente descreve a distribuição de uma quantidade M fixa de bolas em N caixas. Como o nosso interesse é apenas na distribuição de conectividades, então pode-se focar em como cada nó recebe uma nova ligação, de forma que o montante final de ligações que

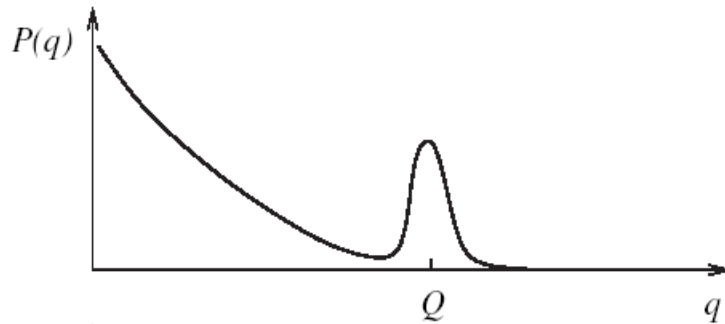


Figura 4.2: Condensação em uma rede de equilíbrio para $\bar{k} > k_c$. A distribuição é reflexo de duas contribuições: uma coincidente com a distribuição no ponto crítico, e a outra devido à condensação das ligações em torno de um conjunto pequeno de vértices com conectividade em torno de um valor Q .

cada nó possuir fornecerá a distribuição desejada. Nesse sentido, o modelo “Backgammon” é formalmente análogo ao processo de construção de redes em equilíbrio. A análise desse modelo foge ao escopo do nosso trabalho, de forma que nos limitaremos a enunciar os resultados cujas deduções podem ser obtidas em [45]. Para $k = k_c$ a distribuição é do tipo lei de potência, para $k < k_c$ a expressão da distribuição é multiplicada por fator exponencial em k e para $k > k_c$ ocorre o fenômeno de condensação. A condensação é um fenômeno em que as ligações condensam-se em torno de alguns poucos nós, i.e., a maioria das ligações é distribuída de acordo com a lei de potência, mas as ligações “em excesso” aglomeram-se em torno de pouco nós. Seguindo um método diferente, Mendes et al. [46] obtiveram a distribuição de conectividades para $k > k_c$, como pode ser visto na figura 4.2.

4.3 Redes “small-world”

Certos sistemas físicos, por exemplo, um cristal, podem ser representados por redes regulares, em que os nós representam os átomos e as ligações representam as interações físicas entre os átomos. Suponha-se um arranjo unidimensional, em que cada átomo se encontre regularmente espaçado em um anel. Dependendo do alcance da interação, cada nó estará ligado a certa quantidade de primeiros vizinhos. Um exemplo de uma rede regular em que cada nó está ligado aos seus $z = 4$ primeiros vizinhos está representado na figura 4.3. Em outro extremo, certos sistemas físicos podem ser representados por redes aleatórias clássicas. Até meados da década de 1990, costumava-se aproximar redes reais como sendo totalmente regulares ou totalmente aleatórias e, a partir de tais estruturas, estudos teóricos eram realizados. Todavia, ao se estudar com mais detalhes certas redes reais, notou-se que a representação em termos de redes situava-se em torno de uma caracterização intermediária entre as redes regulares e as redes aleatórias clássicas. Um primeiro estudo dessa caracterização intermediária foi realizado por Duncan J. Watts e Steven H. Strogatz [4], cujo resultados principais analisaremos.

Iniciemos com uma rede regular como na figura 4.3. No regime $N \gg k \gg \ln N \gg 1$ ($k \gg \ln N$ garante que uma rede aleatória seja conexa [19]) pode-se calcular a distância média \bar{l} entre pares de vértices e o coeficiente de aglomeração médio. A distância média cresce com a quantidade N de nós de acordo com $\bar{l} \sim 2N/z$ e o coeficiente de aglomeração vale $1/2$. Para facilitar a notação, denotemos distância periférica

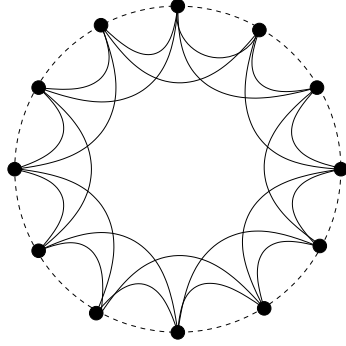


Figura 4.3: Rede regular de dimensão 1 com condições periódicas de fronteira e $z = 4$.

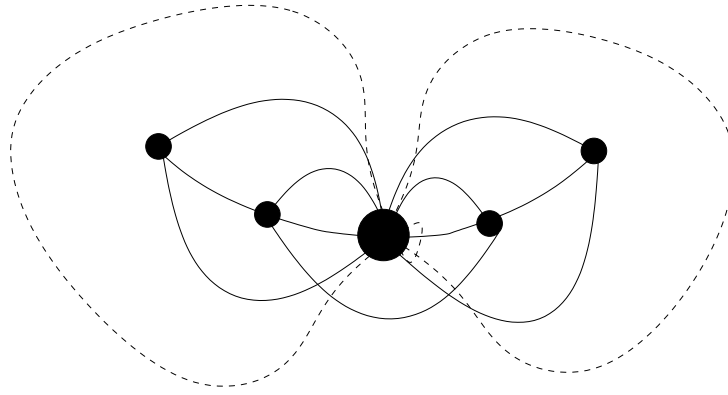


Figura 4.4: Estrutura local de um nó. As linhas tracejadas delimitam os ramos esquerdo e direito.

a distância entre dois nós em que somente as ligações tracejadas são consideradas (essa notação só será usada nessa seção).

Vejam a dedução da expressão da distância média. Suponham-se dois vértices v_0 e v_f distantes. O caminho $\{v_0, \dots, v_i, v_{i+1}, v_f\}$ mais longo é obtido percorrendo as ligações tracejadas na figura 4.3, que fornece a distância periférica. Se os nós estão ligados aos z primeiros vizinhos de maneira regular, como na figura 4.3 em que $z = 4$, então pode-se diminuir o caminho percorrendo as ligações não tracejadas. Com isso, divide-se a distância pelo fator $z/2$. Mas como $N \gg 1$, a maioria dos nós estarão separados pelo caminho mais longo de tal forma que $l_{max} \approx N$. Dividindo-se pelo fator $z/2$, obtém-se a menor distância entre pares de nós.

Já a dedução do coeficiente de aglomeração não é tão imediata. A estrutura local, em torno de um nó, pode ser vista na figura 4.4. Se cada nó possui z vizinhos, então os ramos esquerdo e direito possuem cada um $z/2$ nós. Por sua vez, cada nó pertencente a um desses ramos está ligado aos seus z vizinhos, ou seja, os nós de cada ramo estão todos ligados entre si e a quantidade de ligações existentes dentro de cada ramo é $(z/2(z/2 - 1))/2$. Mas também há ligações entre os ramos esquerdo e direito. Uma análise cuidadosa mostra que o nó que está a uma distância periférica 1 do nó central possui $z/2 - 1$ ligações com nós do outro ramo; o nó cuja distância periférica é 2 possui $z/2 - 2$ e, assim em diante. Dessa forma, a quantidade de ligações entre os dois ramos é $(z/2 - 1) + (z/2 - 2) + \dots + (z/2 - z/2) = (z/2)^2 - \sum_{j=1}^{z/2} j$.

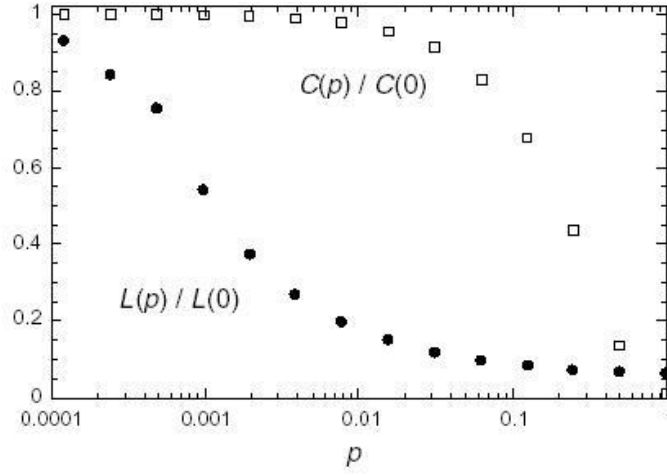


Figura 4.5: Distância média entre dois nós e coeficiente de aglomeração, normalizados por seus valores em $p = 0$, versus p . Esse gráfico caracteriza a transição entre redes regulares e aleatórias [4].

A quantidade total y de ligações entre os z vizinhos do nó central é

$$y = \left[\frac{z}{2} \left(\frac{z}{2} - 1 \right) \right] + \left[\frac{z^2}{4} - \frac{1}{2} \frac{z}{2} (z + 1) \right]. \quad (4.8)$$

No regime considerado, em que $z \gg 1$, o coeficiente médio de aglomeração é $3/4$.

No caso de uma rede aleatória clássica, de acordo com o que vimos no capítulo 2, o coeficiente de aglomeração é $C \sim z/N$, em que z é o número médio de primeiros vizinhos ($z = \bar{k}$) e, se N for grande o suficiente, tem-se $\bar{l} \approx \ln N / \ln z$. Note-se que para N grande o coeficiente de aglomeração da rede aleatória é bem menor do que o da rede regular e a distância média da rede aleatória é bem menor do que o da rede regular. Atribui-se a uma rede, cuja distância média cresce logarithmicamente com o tamanho N da rede, o termo “*small-world*”. As propriedades de mundo pequeno e coeficiente de aglomeração alto são cruciais para o funcionamento de algumas das redes reais e o estudo dessas redes evidenciou a presença de ambas as características.

A fim de estudar a relação entre o aspecto regular e o aleatório, considera-se uma rede regular com N vértices e z ligações por vértice, como na figura 4.3. Com probabilidade p redireciona-se cada ligação para um novo nó escolhido aleatoriamente. Se p é próximo de 0 tem-se a rede regular inicial e se p é próximo de 1 tem-se uma rede aleatória clássica. Com esse procedimento, pode-se obter o coeficiente de aglomeração e as distâncias médias entre os pares de vértices como funções de p : $\bar{l}(p)$ e $C(p)$. Dentro de alguns limites, envolvendo os valores N e z , obtêm-se dois comportamentos distintos: no limite $p \rightarrow 0$ tem-se $\bar{l} \sim N/2z \gg 1$ e $C \sim 3/4$ enquanto que no limite $p \rightarrow 1$ tem-se $\bar{l} \approx \bar{l}_{aleatorio} \sim \ln N / \ln z$ e $C \approx C_{aleatorio} \sim k/N$. A relação entre essas quantidades e p pode ser visualizada no gráfico na figura 4.5.

Note-se que, para pequenos valores de p o coeficiente de aglomeração permanece praticamente o mesmo,

enquanto a distância média diminui rapidamente. O redirecionamento de algumas poucas ligações cria atalhos que diminuem drasticamente as distâncias entre os nós, ao passo que a estrutura local, caracterizada pelo coeficiente de aglomeração, permanece praticamente inalterada. Uma quantidade pequena de atalhos significa uma quantidade da mesma ordem de grandeza que 1. A quantidade de atalhos é $pzN \sim 1$ e, portanto, no limite $N \rightarrow \infty$ haverá uma transição de fase que ocorre no ponto $p = 0$ [47]. Essa transição caracteriza a mudança de uma rede regular, em que as distâncias são grandes para uma rede do tipo mundo pequeno, em que as distâncias são menores. O fato notável é que são necessárias apenas algumas poucas ligações para gerar essa transição.

A rede aleatória clássica é tipicamente conhecida como “*small-world*” devido à relação

$$\bar{l} \sim \ln N. \quad (4.9)$$

Mas, se a rede em questão for uma rede sem-escala, $P(k) \sim k^{-\gamma}$, então a presença dos hubs gera atalhos muito mais eficientes. Nesse caso, se $2 \leq \gamma \leq 3$, a distância média vale [48]

$$\bar{l} \sim \ln \ln N. \quad (4.10)$$

Capítulo 5

Redes fora do equilíbrio

Os sistemas biológicos são sistemas fora do equilíbrio. Uma célula está constantemente trocando material com o meio e sujeita a gradientes de temperatura, de potencial elétrico nas membranas, de concentração de solutos no interior da célula etc.. Além dos fenômenos fora do equilíbrio que ocorrem no interior de uma célula, elementos novos que participam desses processos são criados constantemente a partir de mutações no processo evolucionário das espécies. Durante a evolução, genes novos foram formados, proteínas novas foram sendo produzidas, rotas metabólicas novas se fizeram presentes como parte integrante, enriquecendo os sistemas biológicos mais recentes. Em vista disso, é evidente que as redes biológicas de uma determinada espécie evoluíram. Durante este processo, o número de nós, bem como o padrão das ligações entre esses nós variaram.

Analisando várias redes biológicas atuais, constata-se que possuem uma distribuição de conectividade do tipo lei de potência, isto é, são redes sem-escala. A primeira pergunta que se faz é: que tipo de dinâmica gerou tais redes? Alguns modelos podem ser testados, como um modelo no qual os nós novos são ligados aos antigos aleatoriamente ou um modelo no qual os nós novos escolhem algum antigo de acordo com alguma preferência. Estudaremos alguns modelos de crescimento de redes neste capítulo.

Diferentemente das redes em equilíbrio, teremos uma rede em constante crescimento e haverá uma não-uniformidade na distribuição das ligações: os nós mais velhos sempre possuirão mais ligações. Apesar dos processos que descreveremos conduzirem a um regime fora do equilíbrio, é possível analisar os regimes estacionários de algumas propriedades.

5.1 Crescimento aleatório

Nesse tipo de crescimento, os nós novos ligam-se aos antigos de maneira totalmente aleatória. Tal processo poder ser realizado como se segue. Considerando o tempo discreto, a cada passo temporal um nó novo é acrescentado à rede e liga-se esse nó novo a um outro nó já existente de maneira aleatória. Iniciaremos a partir de uma rede com dois nós possuindo duas ligações entre si. Cada nó será rotulado de acordo com o tempo em que foi criado. Dessa forma, o nó $s = 3$ foi criado no tempo $t = 3$ (os dois primeiros nós possuem rótulos $s = 1, 2$). No tempo t existirão t nós de tal sorte que $\bar{k} = 2L/N = 2t/t = 2$ durante todo o processo. A figura 5.1 ilustra esse processo.

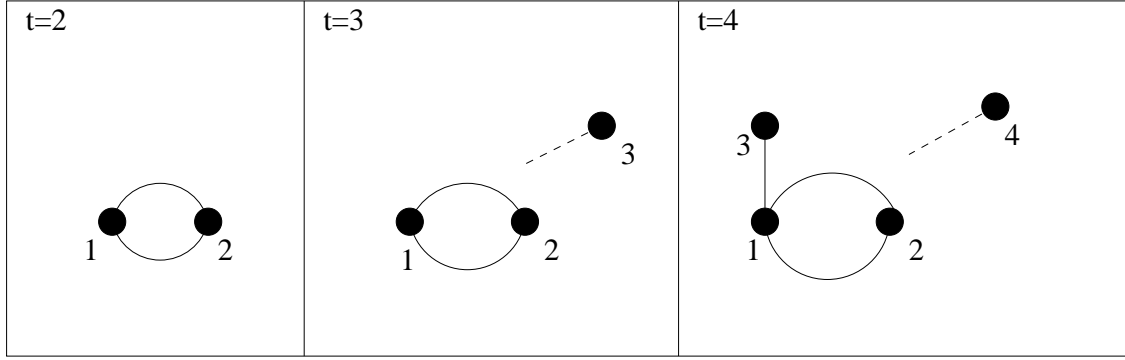


Figura 5.1: Crescimento aleatório. No tempo $t = 2$ inicia-se com os nós 1 e 2. No próximo passo, o nó é acrescentado e ligado aleatoriamente a um dos nós já existentes, que nesse figura é o nó 1. O processo se repete até que a rede tenha N nós e N ligações.

A equação mestra que descreve a evolução de probabilidade é

$$p(k, s, t + 1) = \frac{1}{t}p(k - 1, s, t) + \left(1 - \frac{1}{t}\right)p(k, s, t), \quad (5.1)$$

em que $p(k, s, t)$ é a probabilidade de o nó s ter k ligações no tempo t . Essa expressão é clara. Um nó s no tempo $t + 1$ terá k conexões se ele tiver $k - 1$ conexões no tempo t e receber uma nova ligação, o que ocorre com probabilidade $1/t$. Mas, se o nó s já tiver k conexões no tempo t ele pode permanecer sem ganhar uma nova conexão com probabilidade $(1 - 1/t)$. As condições inicial e de fronteira são $p(k, s, t = 2) = \delta_{k,2}$ para $s = 1, 2$ e $p(k, s = t, t) = \delta_{k,1}$.

A distribuição de conectividades total é

$$P(k, t) = \frac{1}{t} \sum_{s=1}^t p(k, s, t). \quad (5.2)$$

Somando sobre os valores de s na equação 5.1 obtém-se

$$(t + 1)P(k, t + 1) - tP(k, t) = P(k - 1, t) - P(k, t) + \delta_{k,1}. \quad (5.3)$$

No limite do contínuo na variável t tem-se

$$\frac{\partial [tP(k, t)]}{\partial t} = P(k - 1, t) - P(k, t) + \delta_{k,1}. \quad (5.4)$$

A solução estacionária dessa equação é

$$P(k) = 2^{-k}, \quad (5.5)$$

isto é, obtém-se uma distribuição de decaimento rápido. Pode-se calcular a partir da equação 5.1 a conectividade média do nó s , $\bar{k}(s, t) = \sum_{k=1}^{\infty} kp(k, s, t)$ e obtém-se [20], no limite em que s e t são grandes,

$$\bar{k}(s, t > 2) = 1 + \sum_{j=1}^{t-s} \frac{1}{s+j} \cong 1 - \ln(s/t). \quad (5.6)$$

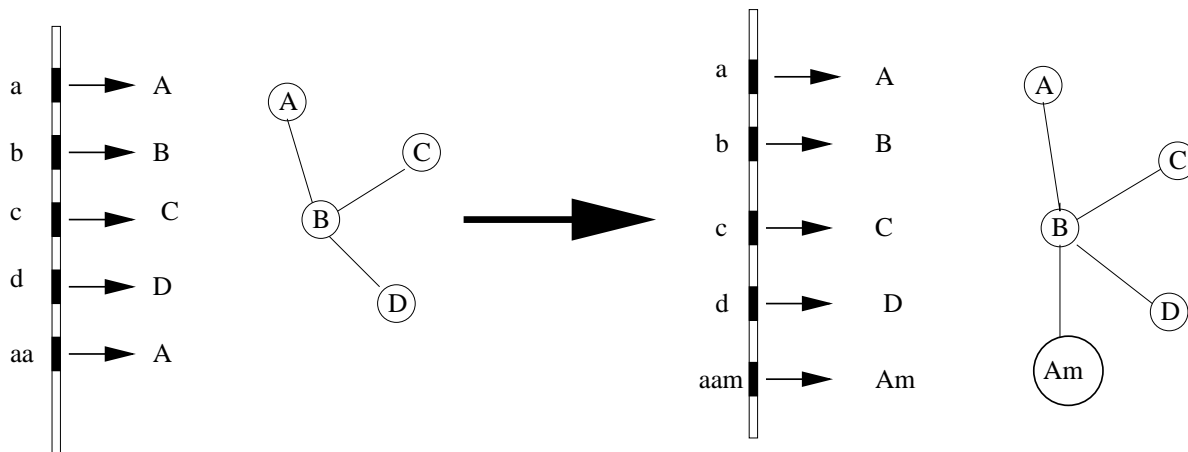


Figura 5.2: Mecanismo de duplicação, mutação e divergência. O gene a possui uma cópia aa e ambos produzem a mesma proteína A . Se uma das cópias sofre mutação, haverá dois genes, a e aam , e poderá haver uma divergência nas funções executadas pelas proteínas, resultando em duas proteínas distintas, A e Am . Se o resultado da mutação não for radical, a nova proteína Am permanece ligada à proteína B .

Note-se que, como era de se esperar, os nós mais velhos, em que $s/t \ll 1$, possuem mais ligações do que os nós mais novos. É importante ressaltar que esse procedimento de construção envolve a composição de um ensemble de redes de tal sorte que a distribuição $p(k, s, t)$ realmente faz sentido.

Alguns outros cálculos são possíveis, como o cálculo da forma da distribuição $p(k, s, t)$. Contudo o importante é constatar que, uma rede em crescimento, na qual os nós novos se ligam aos antigos de maneira aleatória, fornece uma distribuição exponencial e não uma distribuição tipo lei de potência.

5.2 Crescimento de redes sem-escala através de ligações preferenciais

Como já dissemos, as redes biológicas estão em constante crescimento (apesar de a escala de tempo desse crescimento ser muito grande, isto é, a evolução se processa lentamente). Em geral essa evolução ocorre com mutações e seleção natural dos organismos. Um exemplo simples, mas que motivará o modelo de crescimento de redes, é o caso da rede de interação de proteínas. Um fenômeno recorrente na estrutura dos genes é a duplicação. Quando a célula se divide, é necessário fazer uma cópia do seu DNA a fim de que as duas células filhas tenham o mesmo material genético. Às vezes, pode ocorrer que um parte do DNA, conhecida como gene, seja copiada mais de uma vez, o que gerará em uma das células filha genes duplicadas, i.e., genes com duas cópias idênticas na mesma fita de DNA, mas em locais distintos. Como o gene é o mesmo, as proteínas produzidas a partir de ambas as cópias interagem com os mesmos parceiros. Se um desses genes duplicados sofrer uma mutação, as proteínas que antes eram idênticas podem se tornar funcionalmente distintas e manter os mesmos parceiros de interação. Como uma proteína que possui muitas ligações tem maior probabilidade de estar ligada a alguma proteína, cujo gene é duplicado, elas têm mais chances de ganhar uma nova ligação devido a esse processo de duplicação e mutação [49]. A figura 5.2 ilustra esse processo.

Isso sugere um modelo de crescimento de rede em que os nós novos se ligam aos antigos tendo preferência em relação aos que possuam maior conectividade. Esse modelo, conhecido popularmente como “o rico cada vez fica mais rico”, foi introduzido por Albert-Laszlo Barabási e Réka Albert [15]. Nesse trabalho, foi

constatado que, em uma rede em crescimento em que a cada passo temporal um novo nó é acrescentado com ligação preferencial (cada nó novo se liga a algum nó antigo com probabilidade proporcional à conectividade), a distribuição de conectividade é do tipo lei de potência ($P(k) \sim k^{-\gamma}$).

5.2.1 Um modelo com ligações preferenciais

Analisaremos nessa seção um modelo mais geral proposto por Redner et al. [50], em que se considera que, a cada passo temporal, um nó novo é adicionado e é ligado a algum nó antigo com probabilidade determinada por uma função A_k . Essa função depende da quantidade de conexões e sua forma funcional evidentemente determinará o padrão de distribuição das conexões na rede. Nessa seção estudaremos funções A_k tal que $A \sim k^\beta$ no limite $k \rightarrow \infty$. Note que o caso $A_k = k^\beta$ está incluído.

A equação mestra para esse processo é

$$\frac{dN_k}{dt} = A^{-1}[A_{k-1}N_{k-1} - A_k N_k] + \delta_{k,1}, \quad (5.7)$$

em que N_k é quantidade de nós com conectividade k e $A = \sum_{j \geq 1} A_j N_j(t)$. Essa equação se aplica para $k \geq 1$ e define-se $N_0 = 0$. Note que um fator constante multiplicativo na função A_k é irrelevante, pois esse fator é cancelado na equação 5.7, devido à divisão por A . A razão $A_{k-1}/\sum A_k N_k(t)$ é realmente a taxa com que $N_{k-1} \rightarrow N_k$, pois a quantidade $A_{k-1}N_{k-1}/\sum A_k N_k(t)$ é igual a $\Delta N_{k-1}/\Delta t$ fazendo $\Delta t = 1$, pois o caso é discreto.

Para os cálculos que se seguem, será importante calcular alguns momentos da distribuição ($M_n = \sum_{j \geq 1} j^n N_j(t)$). A partir da equação mestra, pode-se multiplicá-la por k^n e somar sobre os valores de k obtendo dessa forma os momentos. Os momentos para $n = 0$ e $n = 1$ são fáceis de serem calculados. Tem-se que $M_0(t) = M_0(0) + t$ é a quantidade de nós e $M_1(t) = M_1(0) + 2t$ é o dobro da quantidade de ligações. Note-se que os dois primeiros momentos são independentes da função A_k .

Como consideraremos funções do tipo $A_k = k^\beta$, tem-se que $A(t) = M_\beta(t)$ e por isso $M_0(t) \leq A(t) \leq M_1(t)$ se $0 \leq \beta \leq 1$. Substituindo as expressões obtida para os momentos nessa desigualdade com $M_0(0) = M_1(0) = 0$ obtém-se

$$t \leq A(t) \leq 2t.$$

Portanto $A(t) = \mu t$ e, analisando a definição de $A(t)$, tem-se que $N_k(t) = tn_k$ é uma possível solução na região $0 \leq \beta \leq 1$. Na realidade consideraremos funções A_k que possuem comportamento linear apenas no regime assintótico. Contudo, desde que não seja escolhido uma função que não se afaste muito do seu comportamento assintótico a desigualdade envolvendo $A(t)$ continua válida.

Substituindo $N_k(t) = tn_k$ e $A(t) = \mu t$ na equação mestra 5.7 tem-se que

$$n_k = \frac{\mu}{A_k} \prod_{j=1}^k \left(1 + \frac{\mu}{A_j}\right)^{-1}, \quad (5.8)$$

e, como $\mu = \sum_{j \geq 1} A_j n_j$, obtém-se a relação implícita

$$\sum_{k=1}^{\infty} \prod_{j=1}^k \left(1 + \frac{\mu}{A_j}\right)^{-1} = 1. \quad (5.9)$$

Usando as equações 5.8 e 5.9 pode-se calcular qualquer distribuição n_k que tenha sido gerada através de uma função, cujo regime assintótico seja do tipo $A \sim k^\beta$ com $0 \leq \beta \leq 1$. O caso $\beta > 1$ deve ser analisado separadamente.

5.2.2 Caso sublinear

No caso sublinear, tem-se que $A_k \sim k^\beta$ no regime assintótico $k \rightarrow \infty$ e $0 < \beta < 1$. O cálculo de n_k é relativamente elaborado [50] e fornece expressões diferentes para cada valor de β nos intervalos $1/(m+1) < \beta < 1/m$ em que m é um número inteiro positivo. O comportamento para todos os valores de m em geral é similar e segue o padrão do caso $m = 1$, em que $1/2 < \beta < 1$,

$$n_k \sim k^{-\beta} \exp \left[-\mu \left(\frac{k^{1-\beta} - 2^{1-\beta}}{1-\beta} \right) \right] \quad (5.10)$$

O valor de μ pode ser obtido numericamente, mas de antemão já sabemos que $1 \leq \mu \leq 2$. Essa pequena variação implica que, no caso sublinear, a distribuição depende fracamente da forma funcional específica de A_k , que se reflete no valor de μ . Também é importante notar que a distribuição de conectividades decai rapidamente no regime assintótico.

5.2.3 Caso linear

O caso linear ocorre quando $A_k \sim k$ no regime assintótico $k \rightarrow \infty$. O procedimento para o cálculo de n_k é o mesmo do caso anterior, fornecendo o comportamento assintótico tipo lei de potência [50]

$$n_k \sim k^{-\gamma}, \quad (5.11)$$

em que $\gamma = 1 + \mu$. Pela definição de $A(t)$ e lembrando que $A(t) = \mu t$, obtém-se diretamente que $\mu = \sum_j A_j n_j$. Os termos dessa soma para os quais j é grande assumem a forma $j^{-\mu+1}$ e para que essa soma convirja deve-se ter $\gamma > 2$. O caso linear fornece distribuições do tipo lei de potência em que o expoente assume valores dependendo da forma funcional de A_k .

Vejamos o caso particular em que $A_k = k$. Substituindo a expressão $N_k(t) = tn_k$ na equação mestra 5.7 tem-se

$$n_k = A^{-1}[A_{k-1}tn_{k-1} - A_ktn_k] = \frac{k-1}{2}n_{k-1} - \frac{k}{2}n_k.$$

Essa equação fornece a relação de recorrência

$$n_k = n_{k-1} \frac{k-1}{k+2},$$

ou seja,

$$n_k = \frac{4}{k(k+1)(k+2)}. \quad (5.12)$$

No limite $k \rightarrow \infty$ a distribuição assume a forma $n_k \sim k^{-3}$. Esse é exatamente o modelo proposto inicialmente por Réka Albert e A.-L. Barabási [15]. O interessante é que o valor do expoente γ da distribuição tem sempre o mesmo valor, mas, se a condição de linearidade for exigida somente no regime assintótico, então o expoente da distribuição pode assumir qualquer valor maior que dois.

Um exemplo de como o expoente da distribuição pode depender da função A_k é o seguinte: considere $A_1 = \alpha$ e $A_k = k$ para $k \geq 2$. Tal função não é idêntica à função $A_k = k$ para todo valor de k , mas é obviamente assintoticamente linear. Nesse caso, após uma série de cálculos [50], o expoente γ assume a forma $\gamma = (3 + \sqrt{1 + 8\alpha})/2$ que mostra claramente a dependência do coeficiente em relação à função A_k .

Ao contrário do que foi visto no caso sublinear, em que a distribuição assintótica possui pouca sensibilidade no que diz respeito à escolha de A_k , se A_k tiver um comportamento assintótico linear, a distribuição assintótica possuirá grande sensibilidade em relação à escolha de A_k . Isso sugere que o valor $\beta = 1$ é um tanto quanto crítico, pois a distribuição depende fortemente de detalhes microscópicos.

5.2.4 Caso superlinear

No caso superlinear tem-se $A_k \sim k^\beta$ no regime assintótico $k \rightarrow \infty$ e $\beta > 1$. Nesse caso discutiremos apenas uma propriedade.

Vamos determinar a probabilidade de que o nó inicial esteja conectado a todos os outros nós. Suponha-se que, durante todo o processo, os nós novos tenham-se ligado ao nó inicial. Após N passos, a probabilidade de um novo nó estar ligado ao nó inicial é $N^\beta/(N + N^\beta)$. A probabilidade de esse padrão de comportamento continuar indefinidamente é

$$P = \prod_{N=1}^{\infty} \frac{1}{1 + N^{1-\beta}}. \quad (5.13)$$

Se $\beta \leq 2$ tem-se $P = 0$. Mas se $\beta > 2$, tem-se $P > 0$ (basta analisar $\ln P$). Com isso vê-se que para $\beta > 2$ o nó inicial pode capturar todas as ligações.

A análise de caso $1 < \beta < 2$ mostra que haverá sempre algum nó que estará ligado a praticamente todos os outros nós.

5.2.5 Correlações entre as conectividades dos nós

Tanto no caso de redes no equilíbrio como fora do equilíbrio, é importante analisar se o processo que levou à construção da rede foi realizado de tal forma que tenham surgido correlações entre os nós. A evolução das redes biológicas pode ter sofrido uma pressão evolucionária tal que algumas correlações tenham surgido a fim de aperfeiçoar algum ganho [11]. Independentemente da origem biológica, o estudo de correlações fornece informações importantes sobre a estrutura da rede.

A correlação, a que nos dedicaremos especialmente, é a correlação entre as conectividades de pares de nós, $P(k, k')$. Essa medida fornece a probabilidade de uma ligação estar conectada a um par de nós com conectividades k e k' . Se não há correlação então

$$P(k, k') = \frac{kP(k)k'P(k')}{[\sum_k kP(k)]^2}. \quad (5.14)$$

Note-se que não se tem $P(k, k') = P(k)P(k')$. Essa medida fornece a probabilidade de se escolher aleatoriamente dois nós que possuem conectividades k e k' . Por outro lado, se a medida está relacionada à ligação e não ao par de nós, pode-se escolher as extremidade de uma ligação como sendo uma das k extremidades ligadas a um nó, com probabilidade $kP(k)$, e como sendo um das k' extremidades ligadas a outro nó, com probabilidade $k'P(k')$.

Estudaremos somente as correlações presente no caso $A_k = k$. Da mesma maneira que na definição de $P(k, t)$, tem-se $P(k, k', t) = N(k, k', t)/N$, em que $N(k, k', t)$ é a quantidade de nós de conectividade k que foram ligados a um nó com conectividade k' . A seguinte equação mestra para a correlação de conectividade é válida

$$M_1 \frac{dN_{kl}}{dt} = [(k-1)N_{k-1,l} - kN_{kl}] + [(l-1)N_{k,l-1} - lN_{kl}] + (l-1)N_{l-1,k}\delta_{k1}. \quad (5.15)$$

No símbolo N_{kl} , para cada par kl , o índice k refere-se ao nó que foi acrescentado por último.

Novamente, sob as hipóteses $M_1 \rightarrow 2t$ e $N_{kl} \rightarrow tn_{kl}$, pode-se obter que [50]

$$n_{kl} = \frac{4(l-1)}{k(k+1)(k+l)(k+l+1)(k+l+2)} + \frac{12(l-1)}{k(k+l-1)(k+l)(k+l+1)(k+l+2)} \quad (5.16)$$

No limite $k, l \rightarrow \infty$ com $y = l/k$ constante tem-se

$$n_{kl} = k^{-4} \frac{4y(y+4)}{(1+y)^4}. \quad (5.17)$$

Fixando-se o valor de k e considerando $k \gg 1$, a distribuição n_{kl} possui um máximo em $y_M \cong 0.372$. Isso implica que um nó cuja conectividade é alta encontra-se ligado com alta probabilidade a um outro nó, cuja conectividade também é alta. Para um dado l com $l \gg 1$ (lembre-se de que na definição de N_{kl} a posição dos índices é relevante), a distribuição n_{kl} atinge seu máximo para $k = 1$. Isso implica que existem nós com conectividade alta que possuem muita ligações com nós que têm $k = 1$.

Quando ambos k e l são grandes obtém-se [50]

$$n_{kl} \rightarrow \begin{cases} 16(l/k^5) & \text{quando } l \ll k \\ 4/(k^2l^2) & \text{quando } l \gg k \end{cases}. \quad (5.18)$$

Se essa rede fora do equilíbrio fosse não correlacionada, deveríamos ter $n_{kl} = n_k n_l$. Mas essa relação não é verdadeira. Basta comparar as equações 5.12 e 5.16. Isso mostra que, em uma rede em crescimento, as correlações surgem espontaneamente. Apesar de o cálculo anterior referir-se a um caso particular, para qualquer função A_k já discutida haverá correlação.

5.3 Organização hierárquica

Algumas redes biológicas apresentam simultaneamente um alto coeficiente de aglomeração e a propriedade de serem redes sem-escala [14, 51]. Além disso, a distribuição do coeficiente de aglomeração em função da conectividade usualmente segue uma lei de potência ($C(k) \sim k^{-\beta}$). Esses são dados empíricos dos quais os modelos tentam se aproximar. A rede aleatória clássica, além de não prover uma distribuição sem-escala, ainda fornece um coeficiente de aglomeração muito pequeno quando o tamanho N da rede é grande ($C \sim N^{-1}$). Uma rede sem-escala, obtida através do crescimento com ligação preferencial (em particular o caso linear), fornece um coeficiente de aglomeração maior, mas a distribuição do coeficiente em função da conectividade é constante ($C(k) \sim cte$). Como conciliar o fato de o coeficiente de aglomeração ser alto, de esse coeficiente ter uma distribuição tipo lei de potência e de se ter a propriedade sem-escala? Um modelo que incorpora essas características é o modelo que gera uma organização hierárquica, em que

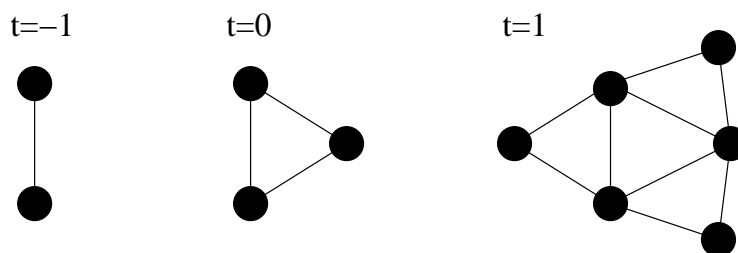


Figura 5.3: Evolução temporal do crescimento da rede determinística. A cada passo uma ligação dá origem a um nó que se liga aos extremos dessa ligação.

pequenos grupos de nós se unem formando grupos cada vez maiores de maneira hierárquica. Analisemos tal modelo.

O modelo de crescimento, exposto na seção anterior, utiliza um procedimento não determinístico, no sentido de que as ligações se fazem de maneira não previsível, apesar de seguir a função de preferência A_k . Utilizemos um procedimento determinístico dessa vez.

O modelo de crescimento inicia em $t = -1$ com dois nós e uma ligação entre eles. A cada passo um novo nó é acrescentado para cada ligação existente. Liga-se esse nó a ambos os terminais da ligação relacionados a ele. O procedimento está ilustrado na figura 5.3.

A cada passo, uma ligação se desdobra em três, de forma que o número total de ligações é $L_t = 3^{t+1}$. A quantidade de nós no tempo t é dada por $N_t = N_{t-1} + L_{t-1} = L_{t-1} + L_{t-2} + N_{t-2}$. Aplicando essa relação recursivamente obtém-se $N_t = 2 + \sum_{k=0}^t L_k$. A somatória nessa expressão pode ser calculada e fornece

$$N_t = \frac{3(3^t - 1)}{2}. \quad (5.19)$$

No tempo t , somente os nós novos possuem $k = 2$ conexões. A quantidade de nós novos é $N_t - N_{t-1} = 3^t$. A quantidade de nós com $k = 4$ é $N_{t-1} - N_{t-2} = 3^{t-1}$. Portanto a quantidade $N(k, t)$ de nós com conectividade $k = 2, 2^2, \dots, 2^{t+1}$ é igual a $3^t, 3^{t-1}, \dots, 3^2, 3$. Para t fixo tem-se, portanto, $N(2^i, t) = 3^{t-i+1}$. Fazendo a mudança de variáveis $k = 2^i$, obtém-se

$$N(k, t) \sim k^{-\ln 3 / \ln 2}. \quad (5.20)$$

O coeficiente de aglomeração pode ser calculado de maneira análoga. Considere-se um nó de referência. Por construção, cada nó novo ligado ao de referência está ligado somente a um dos vizinhos do nó de referência. Dessa forma, somando-se todas as ligações entre os primeiros vizinhos até que o nó de referência possua k primeiros vizinhos, totalizam-se $(k - 1)$ ligações entre os primeiros vizinhos. Portanto $C_t = 2(k - 1)/(k(k - 1))$, ou melhor,

$$C(k) = 2k^{-1}. \quad (5.21)$$

Substituindo a expressão para k em função do tempo, os possíveis coeficientes de aglomeração são $1, 2^{-1}, 2^{-2}, \dots, 2^{-t}$ e a quantidade $m_C(C, t)$ de nós com coeficiente de aglomeração C é $3^t, 3^{t-1}, \dots, 3^2, 3$.

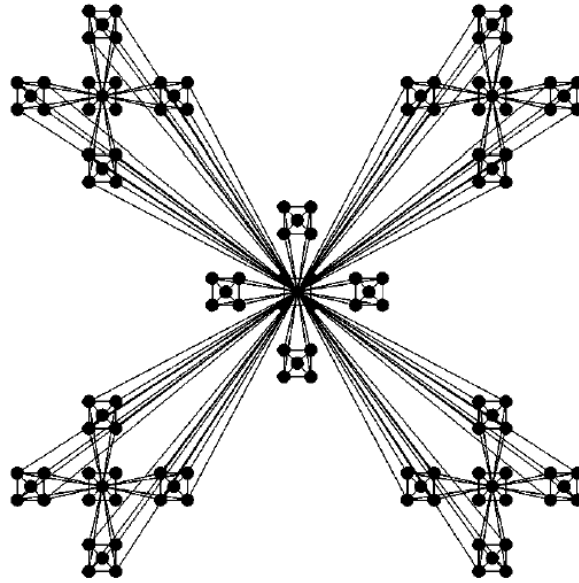


Figura 5.4: Rede hierárquica organizada modularmente.

A expressão fechada é

$$m_C(C, t) \sim C^{\ln 3 / \ln 2}. \quad (5.22)$$

Pode-se calcular o coeficiente de aglomeração médio no limite $t \rightarrow \infty$ obtendo $\bar{C} = 4/5$.

Esta rede com crescimento determinístico pode, portanto, levar a uma distribuição de conectividades do tipo lei de potência, além de apresentar coeficiente de aglomeração médio alto e uma distribuição desses coeficiente do tipo lei de potência.

A fim de incorporar a presença de módulos, isto é, conjuntos de nós que possuem muitas ligações entre os nós do mesmo conjunto e poucas ligações entre nós de conjuntos diferentes, pode-se construir um modelo determinístico que apresenta organização modular, hierárquica e sem-escala [52]. Nesse modelo, as mesmas propriedades estão presentes, mas agora se tem a presença de uma organização modular. A figura 5.4 ilustra esse tipo de rede. A relevância biológica de uma organização hierárquica de módulos é discutida em vários artigos [53, 17].

Capítulo 6

Das redes aos processos fisiológicos

O seqüenciamento completo do genoma humano iniciou uma nova era na biologia conhecida como era pós-genômica [8]. Uma quantidade imensa de dados tornou-se disponível, a partir dos quais foram construídos várias redes biológicas como as citadas no capítulo 3. Contudo há uma diferença crucial entre dados e informação. Embora tenhamos o mapa das redes regulatórias que controlam os processos biológicos, a grande pergunta é:

como capturar as informações contidas na rede e traduzi-las em termos do processo biológico em si?

A fim de responder a essa pergunta, existem vários tratamentos. Pode-se estudar, por exemplo, como a estrutura da rede se organiza em relação às várias funções executadas pela rede. Resultados mostram que as redes biológicas [11, 54, 9, 10] se organizam de maneira modular, isto é, organizam-se de tal forma que existem grupos de nós altamente conectados entre si e pouco conectados com nós pertencentes a outros grupos. Nesse caso, pode-se estudar se os módulos estão correlacionados com funções específicas.

Outra maneira de abordar o problema de capturar as informações é através do estudo de sistemas dinâmicos, cujas definições incorporem a estrutura topológica da rede. Será esse o enfoque que daremos a partir de agora. Nos capítulos finais analisaremos algumas redes regulatórias através de sistemas dinâmicos os quais traduzirão comportamentos e propriedades biológicas.

6.1 Sistemas dinâmicos

Antes de apresentar a definição de sistema dinâmico, vamos definir alguns objetos importantes. Um **vetor de estado** é um objeto que descreve o estado do sistema capturando as características essenciais com que se deseja trabalhar. Para determinar como o sistema evolui no tempo, tem-se uma **função**, ou regra, que determina a seqüência de estados por que passa o sistema. Se considerarmos o espaço Ω dos vetores de estado como sendo o espaço que contém todas as possíveis configurações do sistema, então uma função Γ atuará no espaço dos vetores de estado:

$$\Gamma : \Omega \rightarrow \Omega.$$

Um **sistema dinâmico** é definido como sendo o par (Ω, Γ) .

A caracterização do sistema pode exigir tanto quantidades contínuas, como discretas. Um sistema clássico em que partículas interagem via uma força conservativa pode ser inteiramente caracterizado no

espaço de fase pelas variáveis canônicas posição e momento. Um sistema de N urnas e m bolas distribuídas pode ser caracterizado pela quantidade q_i que representa a quantidade de bolas na urna i . Nesse último exemplo, a quantidade que descreve a configuração do sistema é discreta e o vetor de estado é do tipo (q_1, \dots, q_N) . Nós estudaremos sistemas dinâmicos cujas configurações são do tipo discreto e cuja evolução temporal também é discreta. Poderíamos continuar a enunciar as definições de uma forma geral, mas como o objetivo principal não é o estudo da teoria de sistemas dinâmicos, apresentaremos as definições através de exemplos específicos de sistema dinâmico. O exemplo, que se segue, será o utilizado no estudo das redes regulatórias.

Suponha-se que o sistema em estudo seja representado por uma rede e que o estado do sistema seja determinado pelos estados individuais de cada nó. Suponha-se também que cada nó possa ser caracterizado por dois valores, assumidos no conjunto $\{0, 1\}$. Seja σ_i o valor assumido pelo nó i . Com essas definições o vetor de estado $\vec{\sigma}$ do sistema é da forma $\vec{\sigma} = \{\sigma_1, \dots, \sigma_N\}$ e o espaço de estados é o espaço $\Omega = \{0, 1\}^N$. Resta agora definir a regra Γ que determina a evolução temporal do sistema:

$$\sigma_i(t+1) = \begin{cases} 1, & \text{se } \sum_j a_{ij}\sigma_j(t) > 0 \\ 0, & \text{se } \sum_j a_{ij}\sigma_j(t) < 0 \\ \sigma_i(t), & \text{se } \sum_j a_{ij}\sigma_j(t) = 0 \end{cases} \quad (6.1)$$

em que a_{ij} assume valores inteiros relacionados à ligação direcionada (i, j) (a ligação aponta do nó j para o nó i). Em outras palavras, tem-se que a regra Γ define uma evolução temporal discreta através da relação $\vec{\sigma}(t+1) = \Gamma(\vec{\sigma}(t))$ que está vinculada à topologia da rede através dos coeficientes a_{ij} .

A regra acima define univocamente um vetor de estado $\vec{\sigma}(t)$ para cada vetor $\vec{\sigma}(t-1)$. Se o sistema encontra-se numa dada configuração num tempo t_0 , então para cada tempo posterior (t_0, t_1, \dots, t_T) o sistema possuirá uma configuração determinada pela aplicação recursiva da regra de evolução à configuração inicial $\vec{\sigma}(t_0)$. Essa aplicação recursiva gera uma seqüência de vetores de estado $\{\vec{\sigma}(t_0), \dots, \Gamma^T(\vec{\sigma}(t_0))\}$ que define a **órbita** de uma configuração inicial, em que $\Gamma^1(\vec{\sigma}(t_0)) = \Gamma(\vec{\sigma}(t_0))$, $\Gamma^2(\vec{\sigma}(t_0)) = \Gamma(\Gamma(\vec{\sigma}(t_0)))$ e assim por diante. Cada configuração inicial do sistema, representada por um ponto no espaço dos vetores de estado, possui uma órbita bem definida devido à univocidade da função Γ .

As órbitas no espaço de estados representam a evolução temporal do sistema a partir de cada configuração inicial do mesmo. Dois comportamentos interessantes se destacam. Se uma dada configuração inicial é tal que $\vec{\sigma}(t_0) = \Gamma(\vec{\sigma}(t_0))$ então essa configuração é chamada **ponto fixo**. No caso em que $\vec{\sigma}(t_0) = \Gamma^T(\vec{\sigma}(t_0))$, a órbita a partir de $\vec{\sigma}(t_0)$ é chamada **ciclo de período T** , se T é o menor inteiro com o qual $\vec{\sigma}(t_0)$ é obtido. É evidente que o ponto fixo é um ciclo de período 1. Apenas distinguimos os nomes com a finalidade de evidenciar duas classes de comportamentos, uma em que o estado do sistema é estacionário mas fixo, enquanto na outra classe o sistema é estacionário, mas os estados variam de maneira cíclica.

O que acontece no limite $t \rightarrow \infty$? A princípio a órbita pode percorrer o espaço de estados indo de uma configuração a outra. Mas, se em algum momento a órbita passa por uma configuração que pertença a um ciclo ou ponto fixo então, a partir desse momento, a órbita fica presa ao ciclo ou ponto fixo. As órbitas em espaços mais ricos podem assumir vários comportamentos, mas como o espaço de estados Ω possui uma quantidade finita de estados, a saber, 2^N , somente dois comportamentos $t \rightarrow \infty$ são possíveis: as configurações são atraídas para pontos fixos ou para ciclos.

Os pontos fixos ou ciclos são chamados **atratores** da dinâmica, pois eles atraem as configurações

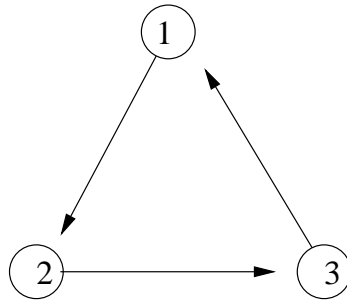


Figura 6.1: Exemplo de uma rede à qual pode-se vincular um sistema dinâmico. As ligações são direcionadas.

iniciais no sentido de que a partir de um certo tempo t_{k^*} a órbita dessas configurações iniciais obedece a $\vec{\sigma}(t_k) = \Gamma^T(\vec{\sigma}(t_k))$ para $t_k > t_{k^*}$. O conjunto das configurações que são atraídas para cada atrator é chamado **bacia de atração** do atrator. Note-se que, na dinâmica, pode haver mais de um atrator e que cada atrator possui a sua própria bacia de atração. Além disso, uma dada configuração inicial não pode pertencer a duas bacias distintas. O tamanho da bacia de atração é definido como a quantidade de configurações iniciais que são atraídas para os atratores.

Outro conceito importante é o de **estabilidade**. Novamente forneceremos a definição desse conceito em termos do sistema dinâmico discreto exemplificado em 6.1. Em verdade faremos uma adaptação do conceito de estabilidade fornecido em [55]. O conceito geral de estabilidade se aplica mal nesse caso, pois o conceito de vizinhança no espaço de estados é trivial, se consideramos o espaço de estados como imerso no Z^N . Uma análise cuidadosa indicaria que todos os atratores são assintoticamente estáveis no sentido fornecido em [55]. Contudo o que se faz é adaptar a noção de estabilidade apropriando-se da noção de que, se a bacia de atração é muito grande, então há uma grande chance de uma órbita perturbada (pode-se perturbar uma órbita mudando o valor de um estado individual de um nó) ainda pertencer à mesma bacia de atração. A definição, portanto, diz respeito mais a uma quantificação de o quanto estável é um atrator. Definamo-la. Seja Σ um atrator, que pode ser tanto um ponto fixo como um ciclo. O atrator Σ é tanto mais estável quanto maior a sua bacia de atração.

Pode-se também classificar um sistema dinâmico quanto à sua **robustez**. A robustez é a medida de insensibilidade do comportamento do sistema em relação a mudanças nos parâmetros que o define. A fim de se avaliar a robustez, é necessário, antes de tudo, definir uma medida para tal, que dependa da situação em questão. Um exemplo de medida de robustez será fornecida na seção 7.3. Quanto às mudanças nos parâmetros, utilizando o exemplo anterior, pode-se fazer variações nos valores do coeficiente a_{ij} que aparece na regra 6.1.

A título de exemplo estudemos um caso trivial. Considere a rede, representada na figura 6.1, vinculada ao um sistema dinâmico discreto governado pela regra 6.1 com $a_{ij} = 1$ para todos as ligações (i, j) . O espaço dos vetores de estado possui 8 estados:

$$\Omega = \{(0, 0, 0), (1, 0, 0), (0, 1, 0), (0, 0, 1), (1, 1, 0), (1, 0, 1), (0, 1, 1), (1, 1, 1)\}.$$

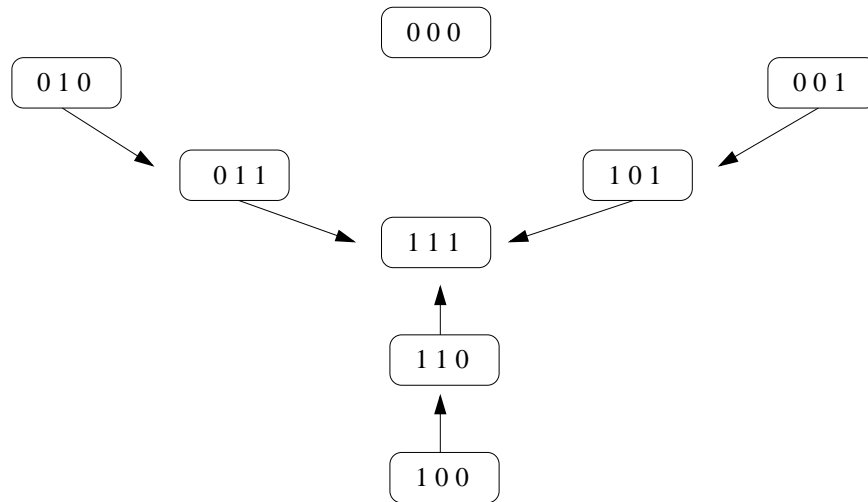


Figura 6.2: Órbitas no espaço dos vetores de estado. Cada nó representa uma configuração $(\sigma_1, \sigma_2, \sigma_3)$. Os atratores da dinâmica são os nós $(0, 0, 0)$ e $(1, 1, 1)$.

No espaço dos estados, pode-se representar cada vetor de estado como um nó e atribuir uma ligação direcionada entre os nós que representam as configurações $\vec{\sigma}$ e $\vec{\sigma}'$ se $\vec{\sigma}' = \Gamma(\vec{\sigma})$ (a ligação é no sentido $\vec{\sigma} \rightarrow \vec{\sigma}'$). Dessa forma, os caminhos presentes nesse grafo representam as órbitas do sistema dinâmico. A simples análise desse exemplo revela a existência de dois pontos fixos, $(1, 1, 1)$ e $(0, 0, 0)$, com as bacias de atração de tamanho 7 e 1 respectivamente (incluímos o ponto fixo na sua própria bacia). A figura 6.2 mostra as órbitas no espaço de estados.

6.1.1 Sistemas dinâmicos contínuos

Em nosso trabalho, estudaremos apenas sistemas dinâmicos discretos. Não só os sistemas dinâmicos discretos, mas também os contínuos são amplamente utilizados na análise de atividades biológicas [56]. Esses sistemas dinâmicos também incorporam em sua definição a topologia de rede que controla o processo modelado.

Nos sistemas dinâmicos contínuos, o vetor de estado é uma grandeza contínua, por exemplo, a concentração $[X_i]$ de uma proteína X . A variação dessa concentração é modelada através de equações diferenciais do tipo

$$\frac{d}{dt}[X_i] = F(X_1, \dots, X_N),$$

em que F é uma função da concentração das demais proteínas.

A análise desses sistemas dinâmicos segue a abordagem formalmente bem estabelecida dos campos vetoriais, bifurcações, atratores etc [56, 55]. Apesar de o nosso estudo utilizar essa mesma abordagem, o fato de ser discreto simplifica sobremaneira a análise, tornando impossível algumas análises mais refinadas. Mas o interessante é que, apesar da simplicidade, propriedades importantes podem ser estudadas.

6.2 Informação biológica

De acordo com o que mencionamos no capítulo 3, as informações necessárias para a execução das atividades biológicas encontram-se armazenadas em uma biblioteca imensa conhecida como genoma. Nessa seção vamos esclarecer melhor os processos básicos que envolvem a transformação dessa informação em atividade biológica propriamente dita¹.

As informações estão codificadas através de um código químico linear contido na molécula de DNA. O DNA é composto de unidades básicas, os nucleotídeos, que se ligam em uma seqüência linear. Os nucleotídeos, por sua vez, podem ser de 4 tipos, dependendo da presença de um entre quatro possíveis componentes (bases nitrogenadas): adenina (A), guanina (G), citosina (C) ou timina (T). Assim o DNA forma uma seqüência linear em que se ordenam 4 unidades distintas, de tal forma que, esse ordenamento possibilita a codificação de informações (AGCAAGT...).

As informações contidas no DNA, na forma do código genético, são traduzidas e utilizadas para construir as proteínas. A proteína é uma molécula formada por unidades básicas conhecidas como aminoácidos, que se ligam linearmente formando uma cadeia. Os aminoácidos existem em 20 tipos diferentes: ácido aspártico (Asp), ácido Glutâmico (Glu), arginina (Arg) etc. Como o idioma do DNA contém 4 letras e o da proteína contém 20 letras, é impossível haver uma tradução biunívoca, do tipo uma adenina se traduz em ácido glutâmico e somente nesse. Nessa tradução, cada conjunto de 3 nucleotídeos, conhecido como **códon**, é traduzido em um aminoácido. Por exemplo, o códon GCA é traduzido no aminoácido alanina. O fato notável é que mais de um códon podem representar o mesmo nucleotídeo. Por exemplo, os códons GCA,GCC,GCG,GCU representam a alanina. A seqüência de aminoácidos, por sua vez, determina as propriedades químicas de ligação das proteínas. A seqüência de aminoácidos garante uma especificidade imensa nas ligações entre as proteínas e as várias outras moléculas o que torna as proteínas agentes especializadíssimos na execução de praticamente todas as atividades biológicas [57]. Um exemplo é a participação das proteínas na atividade metabólica. A célula precisa compor moléculas maiores a partir de subunidades adquiridas via alimentação, bem como decompor moléculas grandes para obter, por exemplo, energia. A síntese e a degradação de moléculas acontecem através de reações químicas que só ocorrem na presença de enzimas. As enzimas são nada mais do que proteínas e o controle da produção das proteínas garante a execução de uma determinada rota metabólica.

A codificação da informação no DNA apresenta uma série de aspectos positivos. As células não precisam manter um estoque de todas as proteínas, pois elas são produzidas somente quando necessárias. Se uma determinada molécula é necessária para a manutenção de uma atividade, mas essa molécula não pode ser adquirida via alimentação, basta recorrer a informação de como produzi-la e isto se encontra no DNA. As informações não se perdem de uma geração para a outra dentro da mesma espécie, pois a informação genética é transmitida integralmente no ato da reprodução. Não menos importante que a manutenção eficiente do *status quo* das atividades celulares é a adaptação a mudanças no meio?

Antes de tratar a questão da adaptação, vejamos um esquema mais formal acerca da idéia de codificação. Na teoria de códigos [58], em geral tem-se a seguinte situação: uma fonte A quer enviar uma informação para um receptor B, mas devido a algumas contingências é necessário codificar a informação. A contingência pode ser o meio pelo qual se transmite a informação. Por exemplo, um astronauta na Lua envia os dados para a Terra através de um código binário, pois o canal que ele tem disponível funciona

¹A revisão teórica apresentada nessa seção foi baseada em [57]

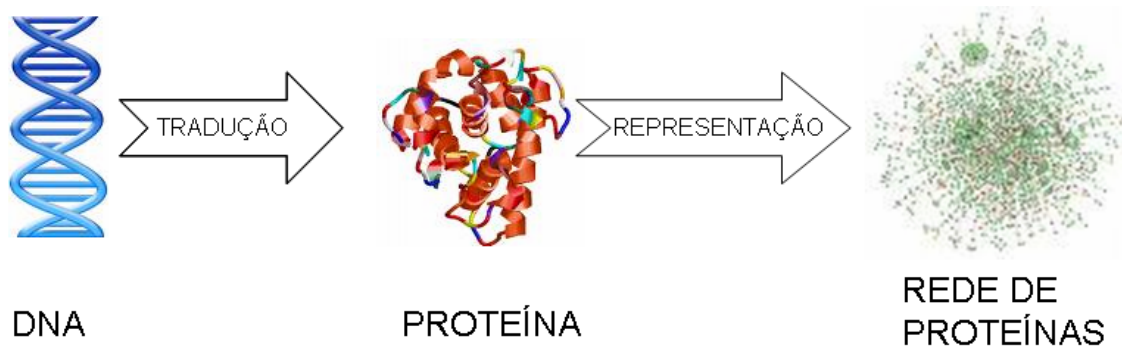


Figura 6.3: Esquema da representação da informação contida no DNA em termos de uma rede. Na primeira etapa traduz-se o gene em proteína. Na segunda etapa, a especificidade das proteínas determina as ligações químicas entre elas que são representadas na rede de interação proteína-proteína. A segunda etapa é realizada pelos métodos descritos na seção 3.1.

com impulsos eletromagnéticos que se adequa ao código binário. É evidente que para que o receptor B entenda a mensagem, é necessário que a tradução da mensagem ocorra. Um conjunto do tipo (110001...) não representa para nós a imagem de uma fotografia tirada na Lua, mas após a tradução desse código em termos de pixels de uma tela de computador, pode-se ver a imagem original. Contudo, se ocorrer algum erro durante a codificação, a imagem reproduzida na tela após a tradução pode ser irreconhecível. Para isso existem algoritmos que tentam corrigir os erros de codificação.

Em relação aos sistemas vivos, surge uma grande diferença no esquema de codificação, tradução e erro. Existe o código, contido no DNA, e existe a tradução do código, em termos de proteínas. Frequentemente, surgem erros na codificação da informação o que leva a construção de uma proteína diferente. Esse é o ponto crucial. No exemplo da transmissão de uma fotografia, se, após a tradução, aparecesse na tela uma imagem indiscernível, diríamos que a informação foi codificada errada e seu conteúdo é inútil. Ao contrário, os sistemas biológicos fazem uso dos erros de codificação para realizarem modificações adaptativas às contingências do meio em que vivem: a nova proteína traduzida pode conferir ao organismo uma maior capacidade de sobrevivência. Os mecanismos que alteram o código genético são basicamente as mutações do DNA.

Juntemos as peças acima, representadas na figura 6.3. O DNA contém informação codificada em termos de 4 signos. Essa informação é traduzida em termos de uma seqüência de aminoácidos que compõe uma proteína. O controle do uso dessa informação bem como as interações necessárias entre as proteínas e outras moléculas podem ser representados através do objeto matemático grafo, formando as redes biológicas descritas no capítulo 3. Essas redes, por ser uma representação das informações genéticas, estão também submetidas à evolução. Vejamos como ocorre o controle do uso da informação.

6.3 Controle da expressão gênica

Suponha-se que, durante o desenvolvimento de um embrião a célula sempre executasse o procedimento de autodestruição, conhecido como apoptose. É evidente que tal ocorrência seria catastrófica para o desenvolvimento da vida. A fim de evitar esse tipo de catástrofe é necessário que as informações contidas nos genes sejam controladas. Não só a fim de evitar catástrofes, mas também a fim de coordenar as várias atividades biológicas, um sistema rigoroso de controle da expressão dos genes é requerido.

6.3.1 Controle transcricional

Uma primeira forma de controle da expressão gênica é o controle transcricional. Vejamos o que vem a ser. Uma molécula cujo envolvimento é crucial nos mecanismos de tradução da informação é o RNA. O RNA é uma molécula semelhante ao DNA, exceto na constituição do tipo de açúcar e no fato de que a base timina é substituída pela uracila (U). Além dessas distinções, os RNA se apresentam somente em uma fita simples. As propriedades químicas de ligação entre os nucleotídeos são as mesmas que no DNA, exceto pela substituição pela uracila.

Os mecanismos de transcrição são extremamente detalhados, por isso, apenas nos ateremos ao mecanismo geral.

A fita dupla do DNA é aberta por uma enzima chamada RNA polimerase (RNAP). Essa enzima, grosso modo, lê a ordem de nucleotídeos no DNA e transcreve essa mesma ordem em uma fita de RNA, só que com os nucleotídeos complementares, isto é, se a fita de DNA contém a seqüência TAGC, a fita de RNA conterá AUUCG. Esse processo é conhecido como **transcrição** e o RNA é conhecido como RNA mensageiro (mRNA)².

Voltando a um problema introduzido na seção anterior, uma fita de DNA contém uma seqüência enorme de nucleotídeos. Como então reconhecer onde começa e termina a receita de uma proteína específica? Primeiramente vamos definir alguns termos. A seqüência de nucleotídeos que contém a receita para a confecção de uma proteína ou RNA é chamada **gene**. O conjunto de genes de uma espécie chama-se **genoma**. Contudo nem toda a fita de DNA é formada de genes. Por exemplo, no caso humano apenas cerca de 27% do DNA corresponde a genes. O início de um determinado gene é marcado na fita por uma seqüência de nucleotídeos específica chamada **promotor** e o término por outra seqüência específica. Dessa forma a RNAP sabe onde se ligar para iniciar a transcrição. Mas uma parte da pergunta persiste: como saber onde está localizado o promotor de um determinado gene especificamente? Em geral, a RNA polimerase liga-se e desliga-se rapidamente de maneira aleatória no DNA, mas quando determinados fatores ligam-se aos promotores dos genes ou a regiões que interagem com promotor, mesmo a longa distância, a RNAP liga-se firmemente ao promotor iniciando a transcrição. Esses fatores são conhecidos como **fatores de transcrição** e podem ser ativadores ou inibidores da transcrição de um determinado gene. Dessa forma, os fatores de transcrição determinam quais proteínas são produzidas e sua presença ou ausência determinam quando se pode produzir tal proteína. Os fatores de transcrição fazem parte, portanto, de um sistema de controle transcricional. A figura 6.4 ilustra esse controle.

Durante o mecanismo de transcrição, a seqüência de nucleotídeos contidas em um gene foi lida e transcrita para uma fita de RNA. Esse processo ocorre no núcleo. Após a transcrição, o mRNA se dirige para uma organela conhecida como ribossomo, situada no citosol, onde ocorre a **tradução** da informação

²Não mencionaremos o fenômeno de *splicing* que antecede a confecção do mRNA

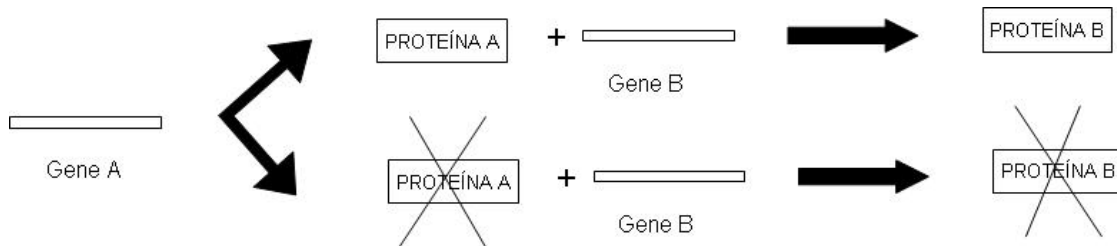


Figura 6.4: Representação do sistema de controle transcricional. Se o gene A não produzir a proteína A, o gene B permanece inativo. Se o gene A produzir a proteína A, esta atua como fator de ativação na transcrição do gene B e a proteína B é produzida. Da mesma maneira que o gene B, o gene A é ativado ou desativado por algum outro fator.

para o idioma das proteínas. Em linhas gerais, a fita de mRNA acopla-se ao ribossomo e este utiliza-se de mecanismos auxiliares para buscar os aminoácidos correspondentes aos códons. Os aminoácidos são ligados em seqüência até que a proteína esteja completamente confeccionada.

6.3.2 Controles pós-transcricionais

Embora não tenhamos detalhado os mecanismos envolvidos nos processos de transcrição e tradução, o fato é que a riqueza imensa de detalhes envolvidos nesses mecanismos permitem a introdução de inúmeros mecanismos de controle da expressão gênica [57] que não os controles transcricionais. A fim de não sobrecarregar o texto de informações que não serão úteis para o desenvolvimento do presente trabalho, apenas nos limitamos a descrever um controle pós-transcricional: o controle da atividade protéica. Mesmo no que diz respeito a esse controle, apenas descreveremos alguns mecanismos que sejam de interesse direto ao presente trabalho. A figura 6.5 ilustra esse processo.

O controle da atividade protéica que abordaremos pode ser visto como um mecanismo de ativação e desativação das proteínas. Mesmo após a confecção de uma proteína, ela pode não se encontrar em uma forma ativa. Num primeiro plano, o que determina as propriedades químicas que garantem a especificidade na ligação de uma proteína ao ligante é a seqüência de aminoácidos. Num segundo plano, a seqüência de aminoácidos determina a conformação espacial da proteína. Contudo, qualquer fator que modifique a conformação de uma proteína pode, a princípio, modificar seu estado de atividade: ativando-a ou desativando-a. O que nos interessa prioritariamente é a modificação de uma proteína via fosforilação.

A **fosforilação** é uma reação química em que um grupo fosfato é adicionado ao aminoácido. Mediante a participação de enzimas essa reação é reversível. As enzimas conhecidas como proteíno-cinase catalisam a reação de fosforilação enquanto enzimas conhecidas como proteíno-fosfatase catalisam a remoção do grupo fosfato. A fosforilação controla a atividade bem como a localização celular de muitas das proteínas das células eucariontes. Como exemplo, algumas proteínas que são produzidas no citosol, só conseguem entrar no núcleo após serem fosforiladas.

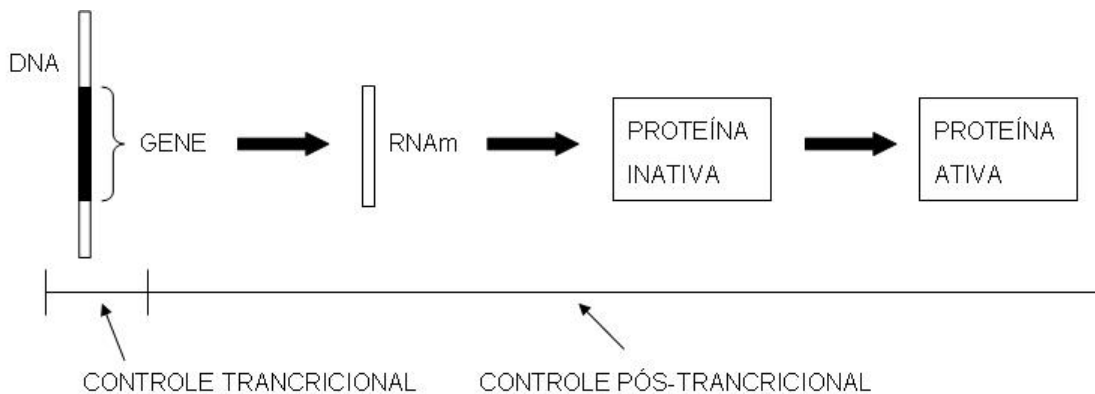


Figura 6.5: Esquema do sistema de controle da expressão gênica. A etapa que compreende a transcrição dos genes pode ser controlada pelo controle transcricional. Todas as demais etapas subsequentes à transcrição são controladas pelos controles pós-transcricionais.

Outro mecanismo de controle é a destruição regulada de proteínas [57]. Uma das maneiras possíveis de se degradar uma proteína consiste na atuação de uma enzima que catalisa a ligação de um marcador, que pode ser tanto uma molécula quanto apenas uma fosforilação à proteína a ser degradada. Alguns complexos enzimáticos exercem a função de degradar proteínas, desde que essas estejam ligadas a um marcador.

6.4 Redes regulatórias e sistemas dinâmicos

Os processos de regulação da expressão gênica envolvem mecanismos transcricionais bem como pós-transcricionais. Enfoquemos primeiramente os controles transcricionais. De acordo com o que vimos no capítulo 3, é possível representar um sistema de controle regulatório de transcrição na forma de uma rede regulatória de transcrição, em que os nós são os genes e os produtos desses genes e as ligações representam ativação ou inibição dos genes por parte dos fatores de transcrição que são, por sua vez, produtos de outros genes.

A atuação dessa rede regulatória de transcrição deve ser capaz de controlar as atividades biológicas de maneira coordenada tanto espacial como temporalmente. Ao envolver o aspecto temporal, é necessário vincular a rede a um sistema dinâmico para que o processo de organização temporal seja capturado na teoria. O primeiro passo para realizar essa vinculação é analisar como os fatores de transcrição atuam no gene e formalizar essa interação.

Considere o pequeno circuito apresentado na figura 6.6. Nesse circuito, os nós X e Y representam fatores de transcrição enquanto o nó Z representa um gene. O símbolo PL mediando a atuação dos fatores de transcrição representa uma porta lógica. Nesse circuito, o fator de transcrição X ativa o fator Y e ambos ativam o gene Z . Mas como se dá a atuação conjunta de ambos os fatores sob o gene Z ? A porta lógica deve representar o que acontece no mecanismo de transcrição. Se ambos os fatores são necessários para que ocorra a transcrição, então, a porta lógica é diferente do caso em que somente um dos fatores é necessário. Uma maneira de formalizar essas noções é a utilização das tradicionais portas

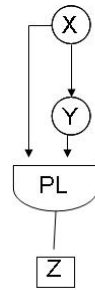


Figura 6.6: Exemplo de um pequeno circuito regulatório “*feed-forward loop*”.

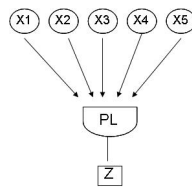


Figura 6.7: Exemplo de um pequeno circuito regulatório “*mult-input*”.

lógicas AND, OR etc. Contudo, num circuito como o da figura 6.7 relações binárias como a porta lógica AND não são suficientes para capturar a atuação dos múltiplos fatores de transcrição. Uma maneira de capturar a atividade de múltiplos fatores de transcrição pode ser realizada considerando que todos os fatores têm influência similar e resultado final da atuação é puramente aditivo. Em outras palavras, se o número de fatores de ativação é maior que o de inibição, então o gene é ativado e em caso contrário é inativado. Quando a quantidade de fatores de inibição e de ativação for a mesma então pode-se considerar a influência dos fatores como neutra e o estado do gene permanece o mesmo, isto é, se ele estiver ativado permanece ativado e se estiver inativado permanece inativado.

A integração de vários circuitos proporcionará um sistema de controle com programação espacial e temporal eficaz no controle dos processos biológicos. Um fato notável, que apenas comentaremos, é que nem todas os tipo de circuitos biológicos se apresentam em proporção igual nas redes regulatórias. Esses circuitos pequenos são conhecidos como motivos (em inglês “*motifs*”) [59, 60]. A figura 6.8 mostra uma série de circuitos que aparecem na levedura *Saccharomyces cerevisiae* [3]. Cada um desses circuitos exerce uma função diferente na programação temporal. Para um estudo minucioso vide [60].

O controle no nível transcricional na maioria dos casos não é suficiente, e a coordenação de uma determinada atividade pode necessitar de um controle pós-transcricional. Ambas as formas de controle podem ser abstraídas na forma de uma rede mais geral, em que se incluem ambos os processos de controle. A essa rede mais geral chama-se **rede regulatória**. Os nós dessa rede podem ser genes, proteínas e RNA produzidos, bem como complexos protéicos. As ligações podem representar (i) a ativação de um gene por um fator de transcrição, (ii) a degradação de uma proteína por complexo de degradação ou (iii) a inibição ou ativação de uma proteína via reação de fosforilação. As mesmas considerações feitas sobre as portas lógicas nas redes regulatórias de transcrição são válidas nessa rede mais geral.

De acordo com o que vimos até agora, a rede regulatória e o sistema dinâmico são os últimos estágios

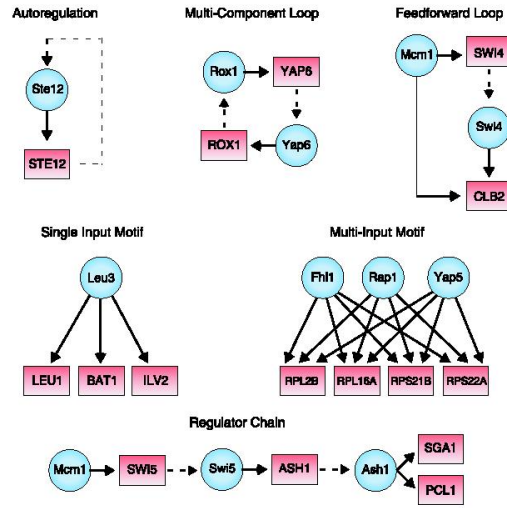


Figura 6.8: Motifs presentes na rede regulatória da *S. cerevisiae* [3].

do processo de transformação da informação biológica. Inicialmente a informação contida nos genes traduz-se na informação contida nas proteínas e RNA. Por sua vez, as informações contidas nas proteínas traduzem-se nas interações que executam as atividades biológicas. O objeto matemático que captura as informações genéticas em termos das interações entre os genes e seus produtos é o objeto grafo e o objeto matemático que captura essas informações em termos de uma evolução temporal é o objeto sistema dinâmico.

De posse da definição de quais portas lógicas medeiam às interações nas redes regulatórias, é possível vincular um sistema dinâmico à rede. No presente trabalho consideraremos apenas a porta lógica do tipo aditiva discutida anteriormente e a definição do sistema dinâmico é a mesma fornecida no exemplo da seção 6.1.

Ao utilizarmos a porta lógica aditiva, estamos fazendo uma aproximação na forma pela qual um gene ou proteína são ativados. Se a ativação de uma proteína Z depende dos fatores de transcrição X_1, \dots, X_N , então a dependência desses fatores pode ser capturada na forma de uma função que envolve as concentrações dos fatores de transcrição $[X_1], \dots, [X_N]$ (nessa notação $[X_i]$ significa a concentração do fator X_i),

$$\text{taxa de produção de } Z = f([X_1], \dots, [X_N]).$$

Essas funções em geral fornecem um valor crítico das concentrações, K_1, \dots, K_N , acima dos quais a expressão da proteína Z é significativa. A princípio, a função f pode ser tão complicada quanto se queira a fim de melhor descrever as interações. No nosso exemplo, a porta lógica aditiva é uma função f que incorpora duas aproximações: (i) desconsidera o papel das concentrações e apenas considera se um fator está presente ou não em sua forma ativa (estado 1 ou 0 do nó); (ii) considera todas as taxas de reações iguais.

Os conceitos de estabilidade e robustez serão importantes na análise dos sistemas dinâmicos vinculados às redes regulatórias. Contudo, ao se estabelecer o significado biológico de tais conceitos, algumas ressalvas

devem ser mencionadas [61]. O vetor de estado deve corresponder a estados com significados biológicos e a evolução temporal de tais estados deve corresponder à execução de funções biológicas. Em vista disso, um atrator estável deve corresponder a um estado biológico que seja realmente de equilíbrio. Suponha que o sistema dinâmico descreva a concentração de uma substância e que a concentração dessa substância deve se manter sempre em torno de um valor para que uma função seja bem realizada. Suponha também que qualquer alteração nas condições a que o organismo esteja submetida acarrete forças restauradoras que culminem na manutenção da concentração em torno desse valor, de forma que esse é um valor de equilíbrio. Nesse caso tem-se uma situação em que a concentração de equilíbrio é estável. Agora suponha outra situação em que haja duas concentrações de equilíbrio, com as quais se pode realizar a mesma função, mas que, sob alterações nas condições, ao invés de o sistema tentar se manter em torno do mesmo valor de equilíbrio, seja mais vantajoso mudar para o outro valor de equilíbrio. Nesse caso tem-se que os valores de equilíbrio não são estáveis, apesar de o sistema tentar globalmente se manter “estável”. O ponto capital na clareza do conceito de estabilidade é a definição de o que seja alterar as condições a que o organismo está submetido. Se essas condições forem incorporadas no sistema dinâmico através de parâmetros que o definem, então uma alteração de parâmetros modifica completamente o sistema dinâmico, isto é, parâmetros diferentes correspondem a sistemas dinâmicos diferentes. Nesse caso os pontos de equilíbrio correspondentes às diferentes concentrações no exemplo dado pertencem a sistemas dinâmicos diferentes e ambos são estáveis dentro do seu próprio sistema dinâmico. Mas se as condições a que o organismo está submetidas está relacionada às condições iniciais de um mesmo sistema dinâmico, então se pode dizer formalmente que os dois pontos de equilíbrio do exemplo não são estáveis.

No caso em que as condições entram como parâmetros do sistema dinâmico, então faz sentido definir a robustez do sistema. Se a medida da robustez for em relação à capacidade de realização de uma função biológica – que pode se vista como uma seqüência de estados regidos pela dinâmica e que possuem significado biológico – então um sistema pode ser robusto mesmo que para cada parâmetro existam mais de um ponto de equilíbrio que não sejam estáveis: desde que a mesma função seja realizada com sucesso mediante a perturbação nos parâmetros e que para cada parâmetro os diferentes pontos de equilíbrio executem a mesma função. Com estas definições, vê-se que um sistema dinâmico pode ser robusto sem ser estável e ser estável sem ser robusto. Em suma, robustez e estabilidade são conceito distintos, embora biologicamente possa se fazer confusão ao se dizer que um sistema é robusto se ele continuar a executar uma função mesmo sob perturbações.

De posse da aliança entre redes e sistemas dinâmicos, estamos prontos para analisar algumas redes regulatórias reais.

6.5 Da biologia celular molecular para a modular

O título dessa seção é a tradução do título do artigo “From Molecular to Modular Cell Biology” [9]. Os autores fazem uma revisão na qual se mostra que as funções celulares são executadas por módulos compostos por várias moléculas que interagem entre si. Uma questão proposta é: como as funções biológicas emergem da complexa rede de interações? É sabido que alguns módulos funcionam isoladamente, mas grande parte deles influenciam-se reciprocamente, seja através de componentes em comum, seja através de uma sinalização externa ao módulo. Além disso, um módulo não precisa exercer uma função específica durante todo o tempo, mas sua função pode ser regulada através de mecanismos outros. Funções podem

emergir em um nível mais alto, em que vários módulos integram-se a fim de executar uma tarefa em comum.

Na concepção estrutural de módulo, módulo é um conjunto de nós altamente conectados entre si e pouco conectados com nós externos ao módulo [62, 63, 64]. Na concepção funcional, módulo é uma entidade discreta, cuja função é separável de outros módulos [9]. A questão da correlação entre ambos os conceitos, isto é, como o módulo estrutural se relaciona ao módulo funcional, é ainda um problema em aberto [10, 11].

Apesar de termos fornecido uma introdução geral à teoria de redes, em que as redes de transcrição, as redes de proteínas, entre outras, foram apresentadas como as redes de todas as transcrições, a rede de todas as interações de proteínas, entre outras, focar-nos-emos no estudo dos módulos funcionais. Veremos como a vinculação entre uma rede, que representa um módulo funcional, e um sistema dinâmico, que modela a evolução temporal das interações, pode nos fornecer propriedades que caracterizam o processo fisiológico.

Capítulo 7

Ciclo celular

Nos capítulos anteriores apresentamos como um processo biológico pode ser representado na forma de uma rede e que a vinculação de um sistema dinâmico à rede poderia capturar a evolução temporal do processo. Nos dois próximos capítulos, estudaremos as redes regulatórias de dois processos biológicos, os quais, através de uma análise via um modelo discreto bastante simples, podem ser bem caracterizados no que diz respeito a algumas propriedades, como robustez e estabilidade. Como seria desejável, o modelo reproduz o processo biológico de maneira fidedigna. A aliança entre a reprodução do processo biológico e a capacidade de capturar propriedades importantes do sistema biológico conferem, ao modelo que será utilizado, a sua validade no estudo de redes regulatórias, sugerindo ser esse um bom modelo para um estudo inicial de redes regulatórias em geral.

7.1 Panorama geral do ciclo celular

Num âmbito mais geral, as células precisam gerar cópias delas mesmas a fim de manterem a existência do sistema vivo ao qual pertencem¹. Se o sistema vivo é um organismo unicelular, o processo de cópia significa a própria reprodução desse organismo. Se o indivíduo é multicelular, o processo de cópia pode significar, por exemplo, a reposição de células mortas em um tecido, ou, então, o desenvolvimento embrionário, no qual a partir de uma única célula, o indivíduo se forma através de várias replicações da célula seguidas de diferenciações quanto ao tipo celular. Independentemente dos detalhes com os quais ocorre o ciclo, um dos objetivos centrais a serem executados é copiar e transferir a informação genética para a próxima geração.

Uma célula é composta de unidades menores chamadas organelas, como os ribossomos, as mitocôndrias, o retículo endoplasmático etc. Em alguns organismos, conhecidos como eucariontes, a célula divide-se em dois compartimentos: o núcleo e o citosol. O material genético encontra-se no núcleo e a maquinaria básica, como as organelas, que executa as informações contidas no DNA situa-se basicamente no citosol. As informações genéticas podem ser divididas em mais de uma fita dupla de DNA. Cada fita dupla de DNA é conhecida como **cromossomo**. A questão da geração de cópias resume-se então a produzir uma cópia idêntica dos cromossomos e a dividir a célula em duas. Esse processo é conhecido como **ciclo celular**.

¹A revisão teórica desse capítulo foi embasada em [57]

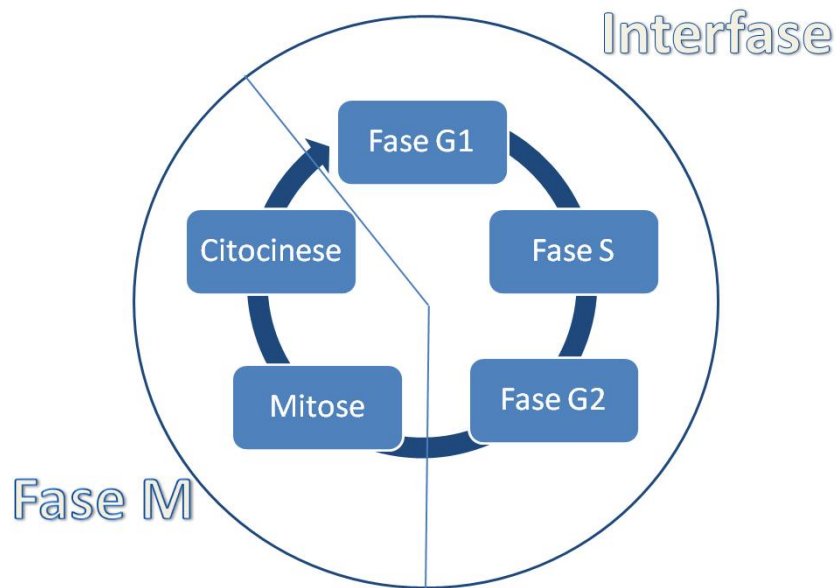


Figura 7.1: Etapas do ciclo celular.

O ciclo celular é dividido em duas grandes etapas: **interfase** e **fase M**. Durante a interfase, a célula aumenta seu tamanho e duplica seu material genético. Na fase M, o material genético duplicado é separado e a célula dividida em duas, de tal forma que cada célula resultante possui uma cópia completa do material genético.

A interfase é dividida em três fases: **G1**, **S** e **G2**. Na fase G1 a célula aumenta a quantidade de seus constituintes a fim de poder gerar duas células filhas completas. Na fase S o material genético é duplicado. Na fase G2 ocorre uma verificação das etapas anteriores, de tal forma que, se todas as etapas anteriores tiverem sido executadas com sucesso, então se pode dar início à fase M.

A fase M é dividida em duas etapas: **mitose** e **citocinese**. Na mitose ocorre a separação do material genético e na citocinese a célula é dividida em duas, cada uma com uma cópia do material genético. A mitose, por sua vez, se divide em prófase, pró-metáfase, metáfase, anáfase e telófase. Essas divisões da mitose caracterizam-se pelo estado de separação dos cromossomos duplicados. A figura 7.1 representa as etapas que compõem o ciclo celular.

Existe uma complexa maquinaria genética e molecular que executa o ciclo, cuja descrição pode ser encontrada em [57]. Em vista dessa imensa complexidade, geralmente elege-se um organismo modelo a fim de se concentrar os esforços de pesquisa. O organismo modelo eucarionte mais simples é a levedura *Saccharomyces cerevisiae*, que é um organismo unicelular pertencente ao reino dos fungos. No presente trabalho apenas nos concentraremos no sistema de controle do ciclo celular da levedura. Não obstante essa restrição em nosso estudo, o ciclo celular da levedura é um bom modelo para os demais eucariontes.

7.2 Sistema de controle do ciclo celular da *S. cerevisiae*

O sistema de controle do ciclo celular é realizado principalmente por eventos de fosforilação que, de acordo com o que vimos no capítulo anterior, é uma forma de controle da atividade protéica em que um grupo

fosfato é adicionado à proteína. A fosforilação é catalisada por enzimas conhecidas por cinases. Essas enzimas controlam o estado de ativação das várias proteínas que executam os procedimentos mecânicos do ciclo: separação dos cromossomos duplicados, alinhamento dos cromossomos, rompimento da membrana nuclear etc..

No caso do ciclo celular, as cinases são de um tipo especial: só funcionam quando ligados a outras proteínas conhecidas como ciclinas. Em vista disso, essas cinases são conhecidas como **cinases dependentes de ciclina (Cdks)**. Existem quatro classes de ciclinas: G1-ciclinas, G1/S-ciclinas, S-ciclinas e M-ciclinas. Cada classe de ciclinas é responsável pelo controle de uma fase, indicada no prefixo da notação. As Cdks se ligam a cada uma dessas ciclinas, formando um complexo ciclina/Cdk. Após a ativação desses complexos, eventos específicos são desencadeados: os eventos da fase G1 são desencadeados pela G1-ciclina; a transição da fase G1 para a S, pela G1/S-ciclina; a fase S pela S-ciclina e a fase M pela M-ciclina. Note-se que as ciclinas e as Cdks não executam propriamente as tarefas do ciclo, mas controlam o estado de atividade das proteínas que executam as tarefas. Dessa forma, a coordenação espaço-temporal do ciclo celular é executada pelo controle da expressão do complexo ciclina/Cdk, o que se dá tanto pelo controle transcricional como pelo pós-transcricional.

A transição para uma fase posterior é submetida à atuação de mecanismos de verificação da boa execução da fase anterior. Apesar da existência desses mecanismos, não os explicaremos, pois, em nosso estudo, assumiremos sempre que as etapas do ciclo são executadas perfeitamente.

De acordo com o que vimos no capítulo anterior, os sistemas de controle transcricionais e pós-transcricionais podem ser representados através de uma rede regulatória. Vamos apresentar inicialmente a rede regulatória do ciclo celular do organismo eucarionte unicelular *S. cerevisiae* de acordo com [5], apesar dessa discussão inicial ser válida para os organismos eucariontes em geral. Agora explicaremos em detalhe o significado de cada nó, bem como o papel que cada ligação cumpre na rede regulatória.

7.2.1 Rede regulatória

A rede proveniente da análise de dezenas de artigos sobre o ciclo celular realizada em [5] está representada na figura 7.2. Os nós representam 3 classes de proteínas: complexos ciclina/Cdk, fatores de transcrição e complexos de degradação. Os nós estão identificados de maneira simplificada. O complexo ciclina/Cdk está representado somente pela ciclina correspondente: por exemplo, o nó Cln3 representa o complexo Cln3/Cdk. Proteínas que exercem funções semelhantes estão representadas pelo mesmo nó: por exemplo, o nó Cln1-2 representa duas proteínas: a Cln1 e a Cln2. Os complexos de degradação estão representados somente por uma das proteínas integrantes do complexo: Cdh1, Cdc20 e Cdc14. Além dessas simplificações, alguns circuitos periféricos foram omitidos. À medida que formos discutindo o significado de cada ligação, daremos mais detalhes dessas simplificações. Nessa rede não estão presentes as ligações responsáveis pelos pontos de verificação do ciclo, de forma que essa rede representa um ciclo celular em que todas as etapas são executadas com sucesso. O único ponto de verificação presente é o referente ao tamanho da célula, mas esse ponto entra como um elemento externo à rede.

As ligações representam interações do tipo ativação ou inibição. Na próxima seção, procederemos ao estudo detalhado dos fenômenos coordenados por essa rede e o entendimento das ativações e inibições se fará claro.

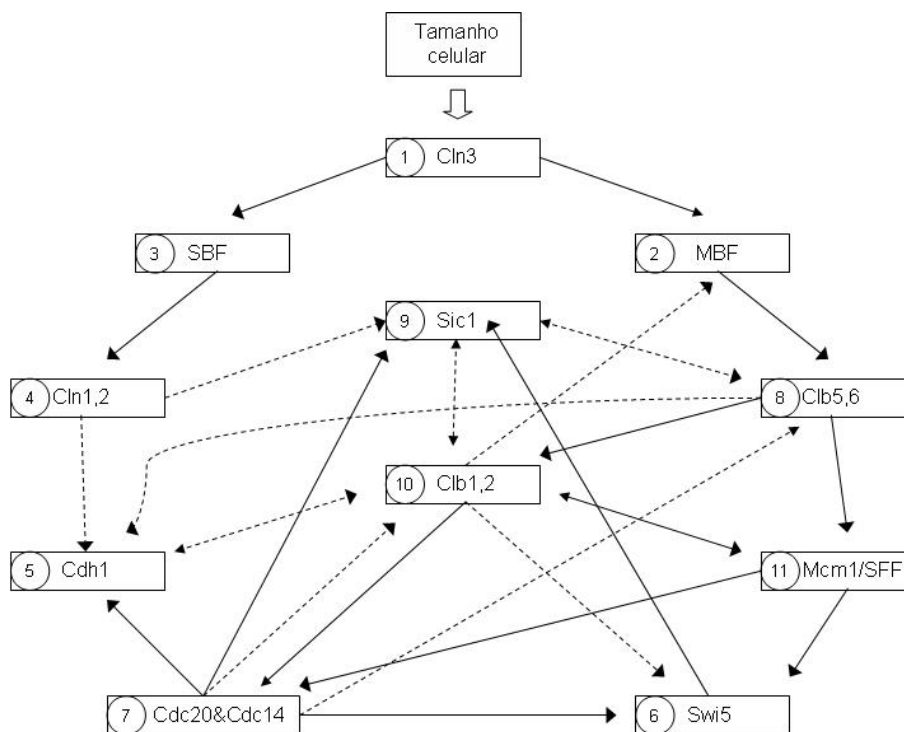


Figura 7.2: Rede regulatória do ciclo celular apresentada em [5]. Os nós estão identificados por um número e pela proteína. As interações de ativação estão representadas pelas setas contínuas e as de inibição, pelas setas tracejadas. Um seta dupla significa que ambos os nós atuam como ativadores ou inibidores recíprocos.

7.2.2 Descrição detalhada das interações de ativação e inibição

Veremos nessa seção o significado biológico das ligações [5, 65, 66, 67, 68, 69, 70, 71, 72, 73] e o que elas representam em termos de reações bioquímicas.

A célula apresenta um estado estacionário que sucede aos eventos da fase M. A célula pode permanecer nesse estado por um período relativamente longo, até que sinais extracelulares e sinais relacionados ao tamanho da célula estejam presentes. Havendo esses sinais, o organismo sai da fase estacionária e coloca-se num estágio próximo ao fim do G1, conhecido como início, quando está pronto para realizar as demais etapas. Na fase estacionária é necessário que os complexos ciclina/Cdk estejam desativados, a fim de que nenhum evento de divisão ocorra. Para que isso ocorra, é necessário que alguns aparatos de degradação e inativação das ciclinas estejam presentes. Vejamos como funcionam esses aparatos presentes durante o estado estacionário.

O aparato de degradação degrada somente a ciclina, mas como a cinase Cdk só funciona quando ligada a uma ciclina, então basta que essas últimas não estejam presentes para que não ocorra os eventos

coordenados pelo complexo. Na realidade, as concentrações das Cdk são praticamente constantes, o que variam são as concentrações das ciclinas. O aparato de degradação funciona de acordo com o processo de marcação descrito na seção 6.3.1. A proteína é marcada por um complexo enzimático e, devido a essa marca, torna-se alvo de um complexo de degradação. Na fase estacionária, o complexo de degradação responsável é conhecido como APC/Cdh1. O complexo enzimático APC, ou complexo promotor da anáfase, é um complexo cuja função varia na dependência de proteínas ligadas à sua estrutura. O nó que representa esse complexo é o nó Cdh1.

O aparato de inativação consiste de uma proteína que, ao se ligar ao complexo ciclina/Cdk, inativa-o devido a mudanças conformacionais. A proteína que exerce essa função no estado estacionário é a Sic1.

A presença dupla do APC/Cdh1 e da Sic1 garantem a repressão completa das ciclinas durante o período estacionário. Vejamos agora como o ciclo celular é iniciado. Alertamos o leitor menos afeito ao estudo de biologia que o restante dessa subseção pode ser enfadonho. O entendimento detalhado das ligações é necessário para que se entenda o significado biológico das ligações, mas se o leitor desejar analisar apenas a parte formal do sistema dinâmico, sintá-se à vontade para saltar o restante dessa subseção.

Fase G1

Consideremos a célula no seu estado estacionário. Quando sinais extracelulares comunicam a viabilidade de uma replicação e quando a célula atinge um tamanho crítico, a ciclina responsável pela fase G1 entra em ação: o complexo Cln3/Cdk acumula-se acima de um nível crítico sendo capaz de atuar em seus alvos. Os alvos desse complexo são proteínas que atuam como fatores de transcrição de genes que também fazem parte da rede regulatória. Esses fatores de transcrição já se encontram presentes, mas na forma inativa aguardando a fosforilação para serem ativados. Os fatores de transcrição estão representados na rede pelos nós SBF e MBF (na realidade esses nós representam mais de um fator de transcrição, o SBF representa as proteínas Swi4 e Swi6, e o nó MBF as proteínas Mbp1 e Swi6). Os fatores de transcrição SBF e MBF são responsáveis pela ativação dos genes responsáveis pela produção de ciclinas necessárias a fases posteriores.

Transição G1/S

Com a ativação do fator de transcrição SBF, as ciclinas Cln1 e Cln2 começam a ser produzidas. Os eventos responsáveis pela saída da fase G1 e entrada na fase S, quando então se dará a replicação do material genético, têm início devido à formação dos complexos Cln1/Cdk e Cln2/Cdk. Nesse exato momento, o outro fator de transcrição ativa a produção das ciclinas responsáveis pela fase S, a Clb5 e a Clb6. Contudo a proteína Sic1, que está presente desde a fase estacionária, liga-se aos complexos Clb5/Cdk e Clb6/Cdk mantendo-os em uma forma inativa. Isso garante que, apesar de tanto as ciclinas responsáveis pela transição G1/S como as responsáveis pela fase S estarem presentes, somente os eventos responsáveis pela transição ocorram.

O complexo Cln1-2/Cdk, além de desencadear eventos ligados à transição, também é responsável pela fosforilação da proteína Sic1 e da Cdh1. A fosforilação também funciona como um marcador para a atuação de um complexo de degradação. No caso da Sic1, o complexo responsável pela degradação é conhecido como SCF e não está representado na rede. A degradação da Sic1 permite o início da fase S, haja vista que os complexos Clb5-6/Cdk permanecem inativos devido a ligação com a Sic1, que não mais está presente.

Fase S

A fase S inicia-se quando os complexos Clb5-6/Cdk estão na forma ativa, o que ocorre graças à

degradação da Sic1. A ativação do Clb5-6/Cdk também ajuda na fosforilação da Cdh1, o que, juntamente com a atuação do Cln1-2/Cdk, garante a inativação do Cdh1.

Transição S/M

No início da transição tem-se o complexo Clb5-6/Cdk ativo e os inibidores Cdh1 e Sic1 inativos. A atuação do Clb5-6/Cdk é um tanto quanto obscura (vide materiais suplementares em [5]). A fim de não gerar um excesso de informação com detalhes desnecessários, apenas diremos que o complexo Clb5-6/Cdk atua como inibidor do Cdh1 e ativador do Clb1-2/Cdk e do Mcm1/SFF (o nó Mcm1/SFF representa o complexo formado pelas proteínas Mcm1, Fkh2, Fdd1). A ativação do Mcm1/SFF bem como a inativação do Cdh1 se dão, entre outros fatores, por fosforilação.

O Mcm1/SFF atua como fator de transcrição da Clb1-2 ativando sua produção. Por sua vez o Mcm1/SFF é também fosforilado pelo Clb1-2/Cdk, o que o ativa. Tem-se na transição um reforço em que o Mcm1/SFF ativa a Clb1-2 e essa ativa o Mcm1/SFF.

Fase M

As quantidades de Clb1-2 e de Mcm1/SFF aumentam simultaneamente. Com a ativação do Mcm1/SFF, as proteínas Swi5 e Cdc20 começam a ser produzidas, pois o Mcm1/SFF é fator de transcrição.

No início da fase M, como o Clb5-6/Cdk e o Mcm1/SFF estão ativos e o Sic1 e o Cdh1 estão inativos, o nível de Clb1-2/Cdk aumenta.

Embora não se tenha certeza, sugere-se que o Clb1-2/Cdk degrade os fatores de transcrição SBF e MBF, o que causa a inativação dos complexos Cln1-2/Cdk e Clb5-6/Cdk. Apesar dessa incerteza, é certo que com o Clb1-2/Cdk ativado os eventos da mitose têm início.

Note-se que os eventos da mitose estão em curso devido ao alto nível de Clb1-2/Cdk. A partir da metade da Fase M torna-se necessário ativar os complexos de degradação da Clb1-2. O Mcm1, como fator de transcrição, ativa a Cdc20. O Clb1-2/Cdk ajuda na fosforilação da Cdc20 ativando-a. A Cdc20 ativa o Cdh1 e o Swi5 através da degradação dos inibidores dessas proteínas. O Swi5 por sua vez atua como fator de transcrição da Sic1 ativando-a.

Com todas as ativações acima tem-se a formação dos complexos de degradação Cdc20/APC e Cdh1/APC bem como a ativação da proteína inibidora Sic1. A atuação desses complexos de degradação e de inibição faz-se de maneira coordenada, de forma que o complexo Cdc20/APC desaparece no final da mitose restando apenas o Cdh1/APC e a Sic1.

A inativação completa das ciclinas garante a célula em um estado estacionário, à espera de sinais para retomar o ciclo celular.

Embora o sistema de controle apresente alguns detalhes cuja explicação é relevante, a colocação de tais informações no texto sobrecarregaria por demais a leitura. A fim de melhor aquilatar o significado das ligações, o material suplementar do artigo base desse capítulo [5] fornece mais explicações e outras referências.

7.2.3 O modelo homogêneo

A rede regulatória que representa o sistema de controle do ciclo celular da *S. cerevisiae* está representada na figura 7.2. Resta-nos, a fim de seguir o programa apresentado no capítulo 6, definir um sistema dinâmico vinculado à topologia da rede. Apresentaremos o modelo de Fangting Li et al. apresentado no artigo “The Yeast Cell Cycle is Robustly Designed” [5].

O modelo dinâmico utilizado em [5] é o mesmo que foi apresentado na seção 6.1. O estado de ativação

e inativação de cada nó é representado por 1 e 0 respectivamente. O estado da célula é representado pelo vetor de estado $\vec{\sigma} = \{\sigma_1, \dots, \sigma_{11}\}$ e o espaço dos vetores de estados contém as 2^{11} possíveis combinações de estados individuais de cada nó.

A evolução temporal do vetor de estado da célula é determinado admitindo-se a hipótese de porta lógica aditiva, de acordo com o que foi visto na seção 6.1. A regra que incorpora a evolução temporal é

$$\sigma_i(t+1) = \begin{cases} 1, & \text{se } \sum_j a_{ij}\sigma_j(t) > 0 \\ 0, & \text{se } \sum_j a_{ij}\sigma_j(t) < 0 \\ \sigma_i(t), & \text{se } \sum_j a_{ij}\sigma_j(t) = 0 \end{cases} \quad (7.1)$$

em que $a_{ij} = 1$ se a ligação $j \rightarrow i$ for de ativação e $a_{ij} = -1$ se a ligação for de inibição. No caso de não haver a ligação $j \rightarrow i$, tem-se $a_{ij} = 0$. A identificação de quais ligações são de ativação ou inibição está representada na figura 7.2. Essa regra define um sistema dinâmico (Γ, Ω) , em que Ω é o espaço dos vetores de estado $\vec{\sigma}$ e Γ é definido da seguinte maneira:

$$\vec{\sigma}(t+1) = \Gamma(\vec{\sigma}(t)) = (F_1(\sum_j a_{1,j}\sigma_j(t)), \dots, F_N(\sum_j a_{N,j}\sigma_j(t))) \quad (7.2)$$

em que $F_i(x)$ é definida por

$$F_i(x) = \begin{cases} 1, & \text{se } x > 0 \\ 0, & \text{se } x < 0 \\ \sigma_i(t), & \text{se } x = 0 \end{cases} \quad (7.3)$$

A dinâmica de cada nó pode ser escrita explicitamente, a saber:

$$\sigma_1(t+1) = 0, \quad (7.4)$$

$$\sigma_2(t+1) = F_2(\sigma_1(t) - \sigma_{10}(t)), \quad (7.5)$$

$$\sigma_3(t+1) = F_3(\sigma_1(t) - \sigma_{10}(t)), \quad (7.6)$$

$$\sigma_4(t+1) = F_4(\sigma_3(t)), \quad (7.7)$$

$$\sigma_5(t+1) = F_5(\sigma_7(t) - \sigma_4(t) - \sigma_8(t) - \sigma_{10}(t)), \quad (7.8)$$

$$\sigma_6(t+1) = F_6(\sigma_7(t) + \sigma_{11}(t) - \sigma_{10}(t)), \quad (7.9)$$

$$\sigma_7(t+1) = F_7(\sigma_{10}(t) + \sigma_{11}(t)), \quad (7.10)$$

$$\sigma_8(t+1) = F_8(\sigma_2(t) - \sigma_7(t) - \sigma_9(t)), \quad (7.11)$$

$$\sigma_9(t+1) = F_9(\sigma_6(t) + \sigma_7(t) - \sigma_4(t) - \sigma_8(t) - \sigma_{10}(t)), \quad (7.12)$$

$$\sigma_{10}(t+1) = F_{10}(\sigma_8(t) + \sigma_{11}(t) - \sigma_5(t) - \sigma_7(t) - \sigma_9(t)), \quad (7.13)$$

$$\sigma_{11}(t+1) = F_{11}(\sigma_8(t) + \sigma_{10}(t)). \quad (7.14)$$

Devido às simplificações contidas na rede, alguns circuitos de degradação estão implícitos. Tal é o caso da Cln1-2. Essa proteína é degradada devido à inativação de seu fator de transcrição. Mas note-se que, se num tempo t ambos Cln1-2 e SBF estão ativados ($\sigma_4(t) = \sigma_3(t) = 1$) e, num tempo posterior, o SBF é inativado ($\sigma_3(t+1) = 0$), então, no tempo $t+2$ teríamos $\sum_j a_{ij}\sigma_j(t) = 0$. Mas isso significaria que o estado da Cln1-2 permaneceria ativo. A fim de incorporar essa degradação implícita, que é necessária para todos os nós que não possuem nenhuma ligação de inibição, assume-se a seguinte regra: se $\sum_j a_{ij}\sigma_j(t) = 0$

BS	Cln3	Mbf	SBF	Cln1-2	Cdh1	Swi5	Cdc20	Clb5-6	Sic1	Clb1-2	Mcm1	Notação
1764	0	0	0	0	1	0	0	0	1	0	0	272
151	0	0	1	1	0	0	0	0	0	0	0	12
109	0	1	0	0	1	0	0	0	1	0	0	274
9	0	0	0	0	0	0	0	0	1	0	0	256
7	0	1	0	0	0	0	0	0	1	0	0	258
7	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	1	0	0	0	0	0	0	16

Tabela 7.1: Tamanho BS das bacias de atração. Cada linha representa os vetores de estado que são atratores da dinâmica.

durante o intervalo $t_0 \leq t \leq t_0 + t_d$ em que o nó i está ativado ($\sigma_i(t) = 1$), então a proteína será degradada no tempo $t_0 + t_d$ ($\sigma_i(t_0 + t_d) = 0$). Como pode ser observado, essa regra vale para os nós Cln3, Cln1-2, Mcm1/SFF e Cdc20&Cdc14. Apesar de o Swi5 possuir uma ligação de ativação, como a sua inativação ocorre quando não mais há Clb1-2, que é o seu inibidor, admite-se que o Swi5 também possui uma inibição implícita. De qualquer forma, essa última inibição foi imposta um tanto quanto arbitrariamente.

A fim de facilitar a notação, como os vetores de estado possuem componentes que assumem os valores 0 ou 1, $\vec{\sigma} = (11000000000)$ por exemplo, pode-se identificá-lo através da seguinte bijeção

$$f : \{0, 1\}^{11} \rightarrow N, \text{ em que } f(\vec{\sigma}) = \sum_{i=0,10} \sigma_i 2^i. \quad (7.15)$$

Nessa notação o vetor $\vec{\sigma} = (11000000000)$ equivale ao estado (I) $I = 3$.

Como todos os coeficientes não nulos são iguais em módulo, esse modelo será chamado modelo homogêneo. Com esse modelo, pode-se estudar o ciclo celular representado por uma seqüência de vetores de estado que evoluem a partir do estado estacionário e analisar algumas propriedades de um sistema biológico, como a robustez.

7.3 A robustez da rede

Na presente seção apresentaremos alguns dos resultados obtidos no artigo [5]. O parâmetro t_d é igual a zero.

Inicialmente verifiquemos os atratores da dinâmica. Seria desejável, pela própria definição de estado estacionário, que o vetor de estado correspondente ao estado estacionário fosse um ponto fixo da dinâmica. De fato, utilizando procedimentos computacionais, nota-se que há 7 pontos fixos e nenhum ciclo. O ponto fixo cuja bacia de atração é a maior é justamente o vetor de estado correspondente ao estado estacionário do ciclo celular, já que Sic1 e o complexo Cdh1 estão ativos e mantêm a inibição estável das ciclinas. A tabela 7.1 mostra os pontos fixos e o tamanho das bacias de atração.

Antes de qualquer conclusão, é preciso verificar se a órbita a partir da configuração correspondente ao estado estacionário corresponde à seqüência de estados por que passa a célula durante o ciclo. A célula sai do estado estacionário se receber o sinal referente ao tamanho e condições extracelulares. Esses sinais se refletem na ativação da Cln3. Por isso, o estado de início do ciclo possui a Cln3, o Cdh1 e o Sic1 ativados e os demais inativados, como se poder ver na linha correspondente ao início na tabela 7.2. A órbita a partir do estado estacionário perturbado está contida na tabela 7.2. Como os estados de ativação das proteínas refletem a fase do ciclo que está sendo executada, uma comparação cuidadosa entre a seqüência

Tempo	Clb3	Mbf	SBF	Clb1-2	Cdh1	Swi5	Cdc20	Clb5-6	Sic1	Clb1-2	Mcm1	Fase	Notação
1	1	0	0	0	1	0	0	0	1	0	0	INÍCIO	273
2	0	1	1	0	1	0	0	0	1	0	0	G1	278
3	0	1	1	1	1	0	0	0	1	0	0	G1	346
4	0	1	1	1	0	0	0	0	0	0	0	G1	14
5	0	1	1	1	0	0	0	1	0	0	0	S	142
6	0	1	1	1	0	0	0	1	0	1	1	G2	1678
7	0	0	0	1	0	0	1	1	0	1	1	M	1736
8	0	0	0	0	0	1	1	0	0	1	1	M	1632
9	0	0	0	0	0	1	1	0	1	1	1	M	1888
10	0	0	0	0	0	1	1	0	1	0	1	M	1376
11	0	0	0	0	1	1	1	0	1	0	0	M	368
12	0	0	0	0	1	1	0	0	1	0	0	G1	304
13	0	0	0	0	1	0	0	0	1	0	0	FASE ESTACIONÁRIA	272

Tabela 7.2: Órbita da configuração inicial correspondente ao estado estacionário perturbado. A coluna tempo corresponde aos passos temporais; as colunas identificadas pelos nós correspondem à evolução temporal dos mesmos; a coluna fase corresponde às fases do ciclo celular; a coluna notação corresponde à notação do vetor de estado de acordo com 7.15.

de estados observados experimentalmente, tal qual descrito na seção anterior, e a seqüência apresentada pela dinâmica mostra que o modelo consegue capturar o evolução temporal do processo biológico de maneira fidedigna. Note-se que o estado estacionário perturbado, como era de se esperar, pertence à bacia de atração do estado estacionário. Logo, tanto o ponto fixo da maior bacia de atração quanto a órbita do estado estacionário perturbado possuem significado biológicos.

O fato de 86% das configurações possíveis serem atraídas para o ponto fixo correspondente ao estado estacionário indica que esse ponto fixo é estável, no sentido definido na seção 6.1. Essa estabilidade possui significado biológico, pois é desejável que, se o sistema for retirado do estado estacionário, ele possa voltar para o estado estacionário. As órbitas no espaço de estado que são atraídas para esse ponto fixo podem ser visualizadas na figura 7.3.

A fim de verificar se a presença de uma bacia de atração grande é fruto da arquitetura particular da rede regulatória ou se é uma característica decorrente de eventos aleatórios, pode-se construir redes aleatórias equivalentes e calcular o tamanho das bacias de atração dos pontos fixos. Uma rede aleatória equivalente é uma rede que possui a mesma quantidade de nós e de ligações, mas a atribuição das ligações é feita de maneira aleatória. Estudando o ensemble dessas redes aleatórias, obteve-se uma quantidade média de 14,28 pontos fixos e uma probabilidade de 10,34 de se obter uma bacia de atração maior que 1764. Isso suporta a conclusão de que a grande bacia de atração não é decorrente de eventos aleatórios.

Não só a estabilidade é desejável, mas também uma robustez em relação à modificação das ligações da rede. A medida que os autores utilizaram para avaliar a robustez diz respeito ao tamanho da maior bacia de atração. Expliquemos essa medida. Considera-se como perturbação das ligações uma das três operações: (i) eliminação de uma ligação, (ii) adição de uma ligação de ativação ou inibição entre nós que não estejam conectados e (iii) troca de uma ligação de ativação por uma de inibição. Avalia-se, para cada modificação da rede, o tamanho BS_p da bacia de atração do estado estacionário e compara-se com o tamanho BS da rede original. A quantidade $|(BS - BS_p)/BS_p|$ é a medida de robustez em relação à perturbação nas ligações da rede. O cálculo dessa quantidade indica que é raro alguma das perturbações eliminar completamente o maior ponto fixo.

O artigo finaliza realizando outras medidas a fim de suportar a afirmativa contida no título do mesmo: a rede do ciclo celular da levedura é desenhada de maneira robusta.

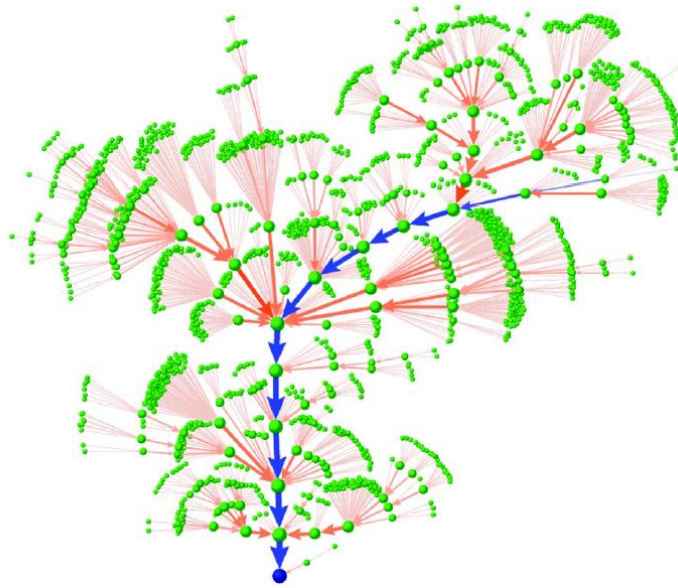


Figura 7.3: Órbitas da maior bacia de atração. O ponto mais inferior na figura corresponde ao estado estacionário 272. [5]

7.4 Intermissão

Até o presente momento foi realizada uma revisão de alguns tópicos em teoria de redes e em biologia. Nenhum dos resultados apresentados foram de nossa autoria. A partir da seção que se segue até o final da dissertação, todos os resultados foram de nossa inteira autoria.

7.5 Aspectos do modelo homogêneo

7.5.1 Ligações duvidosas e circuitos ausentes

Algumas considerações se fazem necessárias sobre as simplificações envolvidas na construção da rede regulatória. No artigo original [5], as ligações $Clb5-6 \rightarrow Mcm1/SFF$ e $Clb5-6 \rightarrow Clb1-2$ são tidas como duvidosas. Essa dúvida diz respeito mesmo às experiências biológicas e não a meras conjecturas sobre o modelo de sistema dinâmico adotado. O que acontece com a dinâmica se essas ligações duvidosas forem retiradas? Obtêm-se 11 pontos fixos ao invés de 7. O estado estacionário da rede completa, que na notação é o 272, ainda possui a maior bacia de atração ($BS=895$). A órbita a partir do estado estacionário perturbado evolui até estacionar no estado correspondente à fase S ($I=142$).

Outro aspecto da simplificação consiste na omissão de um circuito envolvendo o nó $Clb1-2$. Ao final da fase G2 há o aumento da $Clb1-2$, que se dá geralmente pelo aumento na transcrição do gene $Clb1-2$. Simultaneamente a ciclina se liga ao Cdk ocasionando um aumento gradual do complexo $Clb1-2/Cdk$. Esse complexo, para ser ativado, precisa ser fosforilado. Quem realiza esse papel é a proteína CAK. Contudo imediatamente a essa primeira fosforilação, o complexo é preso a um estado inativo por mais uma fosforilação catalisada pela proteína Wee1. Ao final da fase G2 há um grande estoque de $Clb1-2/Cdk$ inativo devido à adição de dois grupos fosfato. A ativação do $Clb1-2/Cdk$ é feita pela remoção

do fosfato inibidor, o que é realizado pela proteína Cdc25. Esse mecanismo promove a rápida ativação de Clb1-2/Cdk haja vista que a reação de fosforilação ocorre numa taxa muito maior do que a taxa de transcrição a partir do gene. Inserindo as interações entre as proteínas Cdc25 e Wee1 e as já presentes na rede original [74] pode-se construir uma nova rede com 13 nós. A órbita a partir do estado estacionário perturbado é similar à da rede original e o ponto fixo correspondente ao estado estacionário é o maior da dinâmica.

Com o que foi explicado, as duas ligações duvidosas são essenciais no modelo enquanto o circuito envolvendo a Cdc25 e a Wee1 não o são.

7.5.2 Variações no tempo de degradação implícita

Os nós representando as proteínas Cln3, Cln1-2, Cdc20&Cdc14, Mcm1/SFF e Swi5 possuem um mecanismo de degradação implícita controlado pelo parâmetro t_d .

Se a degradação implícita estiver ausente ($t_d = \infty$), tem-se 14 pontos fixos e $BS(272) = 1$. A maior bacia de atração é $BS(304) = 1383$. O ponto fixo 304 corresponde ao ponto fixo 272 mais o Swi5 ativado. Se o ponto fixo $I = 304$ é perturbado para $I' = 305$, o sistema evolui em 5 passos até o ponto fixo 302. Não há nenhuma órbita que corresponda ao ciclo celular.

Vejam agora o comportamento do sistema quando se varia o parâmetro t_d .

Começaremos com a degradação implícita do nó Cln3. O tamanho da bacia de atração de cada um dos 7 pontos fixos muda levemente, se variarmos t_d entre os valores 0 e 4. A partir de $t_d = 4$ o tamanho das bacias permanece o mesmo. Acima de $t_d = 6$, o nó Cln3 permanece ligado por um período de tempo longo no ciclo celular, o que acrescenta, na órbita correspondente ao ciclo celular, a repetição do estado 1775, i.e., o sistema permanece nesse estado durante $(t_d - 6)$ passos.

No modelo homogêneo, o nó Cln1-2 é degradado no tempo 8. O tamanho da bacia de atração de cada ponto fixo varia um pouco até $t_d = 6$. Para $t_d > 6$ o tamanho das bacias permanece constante. O tamanho da órbita do ciclo celular aumenta em uma unidade para cada unidade aumentada no tempo de degradação: cada vez que se aumenta o tempo de degradação adiciona-se o estado 1640 após o 1736 de maneira repetitiva. Isso implica que o sistema permanece t_d passos no estado 1640.

Os demais nós seguem o padrão do Cln1-2. Podemos concluir que a dinâmica dos nós Cln3, Cln1-2, Mcm1/SFF, Cdc20&Cdc14 e Swi5 basicamente não mudam, se o tempo de degradação implícito t_d for maior que 0, exceto pelo fato de que eles permanecem ligados por mais tempo, ocasionando uma repetição de estados na órbita. De maneira mais geral, pode-se concluir que a variação no tempo de degradação não altera significativamente a órbita correspondente ao ciclo celular. Um tempo de degradação maior apenas insere estados repetitivos na órbita.

Note-se que os passos temporais não possuem correspondência com as durações das fases do ciclo celular. O modelo captura apenas a seqüência temporal, pois todas as ligações são aproximadas como reações de mesma taxa e limiares de ativação. Esses resultados mostram que o sistema dinâmico é robusto em relação à variações no parâmetro t_d .

7.6 O modelo não-homogêneo

O artigo que foi estudado [5] apresenta a análise de um modelo em que os coeficientes a_{ij} são todos iguais em módulo, i.e., um modelo homogêneo. Os autores afirmam que a dinâmica não sofre alteração se for imposta certa não-homogeneidade do tipo $a_{ij} = a_p$ para as ligações de ativação e $a_{ij} = -a_n$ para as ligações de inibição. Na presente seção, abordaremos o problema de assumir a não-homogeneidade mais geral possível.

Os coeficientes a_{ij} estão relacionados a como o estado de um nó é determinado pela ação conjunta de vários nós: a_{ij} confere um peso para a ligação $j \rightarrow i$. Assumir homogeneidade significa que todos os nós têm a mesma capacidade de ativar ou inibir. Como exemplo, analisemos o nó Cdh1. Suponha-se que tivéssemos $a_{5,7} = 4$ (que se refere à ligação Cdc20&Cdc14 \rightarrow Cdh1) e $a_{ij} = 0$ para as demais ligações. Nesse caso, bastaria o nó Cdc20&Cdc14 estar ativado no tempo t para o Cdh1 estar ativado no tempo $t + 1$. Isso implicaria que o nó Cdc20&Cdc13 possui uma capacidade de ativar que é prioritária em relação às demais ligações. Prioridades distintas nas ligações podem ser incorporadas em um modelo não-homogêneo, em que os coeficientes a_{ij} podem assumir valores irrestritos. Em outras palavras, esses coeficientes determinam a porta lógica (vide 6.1).

Modificar os valores dos coeficientes significa alterar a definição do sistema dinâmico. Mas será que a mudança desses valores realmente define novos sistemas dinâmicos? Para responder essa questão precisamos observar dois pontos. O primeiro é que a regra 7.1 considera somente se a soma $\sum_j a_{ij}$ é negativa, positiva ou zero. Isso implica que vários conjuntos diferentes de coeficientes representarão o mesmo sistema dinâmico. O segundo ponto é que dois sistemas dinâmicos A e B são iguais se, e somente se, $\vec{\sigma}_A(t+1) = \vec{\sigma}_B(t+1)$ para todos os possíveis $\vec{\sigma}(t)$ no espaço de estados. Essas duas considerações são suficientes para se definir o conjunto de coeficiente que descrevem todas as possíveis não-homogeneidades, excluindo as possíveis redundâncias. Formalizemos essas considerações.

De acordo com o que vimos na seção 6.1, um sistema dinâmico constitui-se de uma função $\Gamma : \Omega \rightarrow \Omega$, em que Ω é o espaço de estados. Seja U um subconjunto de Ω . Se $\Gamma(U) = \Gamma'(U)$ para todo subconjunto U , então os sistemas dinâmicos (Γ, Ω) e (Γ', Ω) são idênticos. No modelo não homogêneo, o sistema dinâmico é definido pelo conjunto $C = \{a_{i,j}; 0 \leq i \leq 11, 0 \leq j \leq 11\}$. É possível que uma classe² $[C]$, contendo mais de um conjunto de coeficientes C , defina o mesmo sistema dinâmico? Se isso ocorrer, então as possíveis não-homogeneidades não redundantes são obtidas ao se considerar apenas um representante de cada classe $[C]$.

Antes de prosseguir, modifiquemos a notação dos coeficientes. Para as ligações de ativação utilizemos $a(i, j)$, em que $a(i, j) = a_{ij}$. Para as ligações de inibição, utilizemos $b(i, j)$, em que $b(i, j) = -a_{ij}$. Além disso, passaremos a identificar os nós somente pelos números da figura 7.2.

A regra de evolução temporal pode ser escrita de maneira explícita. Para os nós que só possuem ligações positivas

$$\sigma_1(t+1) = 0, \quad (7.16)$$

$$\sigma_4(t+1) = F_4(a(4, 3)\sigma_3(t)), \quad (7.17)$$

$$\sigma_7(t+1) = F_7(a(7, 10)\sigma_{10}(t) + a(7, 11)\sigma_{11}(t)), \quad (7.18)$$

$$\sigma_{11}(t+1) = F_{11}(a(11, 8)\sigma_8(t) + a(11, 10)\sigma_{10}(t)), \quad (7.19)$$

²Estamos considerando classe como um conjunto de conjuntos. Em particular essa é uma classe de equivalência.

para os que possuem somente duas ligações

$$\sigma_2(t+1) = F_2(a(2,1)\sigma_1(t) - b(2,10)\sigma_{10}(t)), \quad (7.20)$$

$$\sigma_3(t+1) = F_3(a(3,1)\sigma_1(t) - b(3,10)\sigma_{10}(t)), \quad (7.21)$$

para os que possuem três

$$\sigma_6(t+1) = F_6(a(6,7)\sigma_7(t) + a(6,11)\sigma_{11}(t) - b(6,10)\sigma_{10}(t)), \quad (7.22)$$

$$\sigma_8(t+1) = F_8(a(8,2)\sigma_2(t) - b(8,7)\sigma_7(t) - b(8,9)\sigma_9(t)), \quad (7.23)$$

para o que possui quatro

$$\sigma_5(t+1) = F_5(a(5,7)\sigma_7(t) - b(5,4)\sigma_4(t) - b(5,8)\sigma_8(t) - b(9,10)\sigma_{10}(t)), \quad (7.24)$$

e para os que possuem cinco

$$\sigma_9(t+1) = F_9(a(9,6)\sigma_6(t) + a(9,7)\sigma_7(t) - b(9,4)\sigma_4(t) - b(9,8)\sigma_8(t) - b(9,10)\sigma_{10}(t)), \quad (7.25)$$

$$\sigma_{10}(t+1) = F_{10}(a(10,8)\sigma_8(t) + a(10,11)\sigma_{11}(t) - b(10,5)\sigma_5(t) - b(10,7)\sigma_7(t) - b(10,9)\sigma_9(t)). \quad (7.26)$$

Ilustraremos o procedimento de determinação do menor conjunto de coeficientes para o caso da equação 7.18. Note-se que os termos, na parte direita da equação, possuem o mesmo sinal. Para quaisquer valores assumidos pelo par $C = (a(7,10), a(7,11))$, todas as possíveis combinações de estados iniciais $(\sigma_{10}(t), \sigma_{11}(t))$, a saber $D = \{(0,0), (1,0), (0,1), (1,1)\}$, fornecem o mesmo caso (positivo, negativo ou 0) para a soma $\sum_j a_{ij}\sigma_j(t)$, a saber, $S = \{0, +, +, +\}$ na mesma ordem que no conjunto D. Como é apenas o sinal dessa soma que determina o estado $\sigma_7(t+1)$, tem-se que todos os possíveis valores do par C fornecem o mesmo mapeamento entre D e S. Se dois sistemas dinâmicos possuem o mesmo mapeamento entre D e S, então esses sistemas dinâmicos são idênticos entre si (uma análise mais cuidadosa mostra que essa condição, apesar de se referir à dinâmica de um nó, é suficiente para determinar a identidade entre os dois sistemas dinâmicos). Assume-se, portanto, qualquer valor do par C, sendo mais simples assumir o par $P = (1,1)$. Nesse caso, tem-se uma única classe [C] que define o sistema dinâmico, sendo constituída por todos os pares de números inteiros. Note-se que é desnecessário assumir valores não inteiros para os coeficientes.

Para as demais equações, segue-se o mesmo procedimento, exceto pelo fato de que o levantamento das classes e a escolha de um representante de cada classe são feitos computacionalmente. As quatro primeiras equações admitem uma única classe, em que os coeficientes são todos iguais a 1. As equações 7.20 e 7.21 admitem 3 classes, cujos representantes são $\{(1,1), (2,1), (1,2)\}$. As equações 7.22 e 7.23 admitem 11 classes, cujos representantes, por exemplo da equação 7.22, são

$$\{(1,1,1), (1,1,2), (1,1,3), (1,2,1), (1,2,2), (1,3,2), (2,1,1), (2,1,2), (2,2,1), (2,2,3), (3,1,2)\}.$$

As equações 7.24 e refr7 admitem 67 classes e a equação 7.26 admite 3265. Portanto o conjunto mínimo de coeficientes considerando todas as interações do modelo contém $3^2 \cdot 11^2 \cdot 67 \cdot 3265^2 = 7,779019 \times 10^{11}$

elementos. Essa enorme quantidade implica que será preciso utilizar simulações numéricas para estudar o modelo não homogêneo.

7.6.1 Resultados e análises

Em vista da quantidade imensa de casos de não-homogeneidade, analisar todos seria computacionalmente inviável. Dentre as $7,779019 \times 10^{11}$ possibilidades de conjuntos de coeficientes não redundantes, buscou-se fazer uma escolha aleatória e uniforme nesse espaço de conjuntos de coeficientes. Para isso, estudamos uma amostra cujo tamanho é 10^6 . Note-se que essa é uma amostra de sistemas dinâmicos. Adotamos dois procedimentos de análise. O primeiro consiste em analisar os atratores, bem como o tamanho das bacias de atração. O segundo consiste em analisar as órbitas a partir do estado estacionário perturbado.

Para cada conjunto de coeficientes, estudamos a evolução de cada um das 2048 configurações iniciais e determinamos os pontos fixos (ciclo de período 2 e 3 são raros) e suas bacias de atrações. Em nossas simulações, achamos que o menor tamanho da bacia de atração do estado 272 (estado correspondente ao estado estacionário do ciclo celular) é $BS_{min}(272) = 11$. Essa bacia mínima aparece em 0.23% dos casos, quando então se tem 5 pontos fixos ao invés dos 7 achados para o modelo homogêneo. O tamanho médio da bacia de atração do ponto fixo 272 é $BS_{med}(272) = 1622,9$, que é menor do que a bacia no caso homogêneo, $BS_{homog}(272) = 1764$. Contudo 65,3% da amostra possui $BS(272) > 1764$.

A seguir, procedemos ao estudo da órbita do estado estacionário perturbado, i.e., a órbita do estado 273. Encontramos 336 órbitas distintas. A órbita

$$Orb_{freq} = \{273, 278, 286, 14, 142, 1678, 1736, 1632, 1904, 1392, 386, 304, 272\}$$

com 12 passos temporais é a mais freqüente, aparecendo em 10,9% da amostra. Já a órbita que ocorre no modelo homogêneo aparece somente em 3,6% dos casos. Agora considere somente os exemplares da amostra que possuem a órbita Orb_{freq} . Nessa sub-amostra, o tamanho médio das bacias de atração do estado 272 é $BS_{freq}(272) = 1866,7$. Note-se que esse valor médio é maior do que o valor $BS_{homog}(272)$. A única diferença entre Orb_{freq} e a órbita do caso homogêneo é que na fase M o nó Cdh1 é ativado no tempo 9 no caso da Orb_{freq} e no tempo 11 no caso homogêneo.

A saída da fase M é controlada pelo complexo APC, que é ativado principalmente pela Cdc20 e Cdh1. A proteína Cdc20 ativa o APC que inicia a degradação da Clb1-2 durante a transição metáfase-anáfase. A segunda etapa de degradação da Clb1-2 ocorre pela ligação da Cdh1 ao APC durante a telófase. Portanto, a degradação da Clb1-2 necessita de uma ação seqüencial [75]: (i) a ligação do Cdc20 e (ii) a ligação do Cdh1, ambos ao APC. Embora a degradação pelo Cdc20/APC é suficiente para a saída da fase M, a degradação pelo Cdh1/APC é essencial para manter a fase G1 estável [76]. No modelo homogêneo, o Cdh1 é ativado simultaneamente à desativação da Clb1-2, implicando que esse segundo mecanismo de degradação não está presente. Ao contrário, esse segundo mecanismo está presente na órbita Orb_{freq} , haja visto que o Cdh1 e o Clb1-2 permanecem ambos ligados durante dois passos de tempo. Com isso concluímos que a ativação do Cdh1 sendo feita mais cedo é mais coerente com o mecanismo de degradação da Clb1-2.

Os resultados nos permitem concluir que, em relação ao tamanho da bacia de atração do estado estacionário 272, o sistema dinâmico é robusto sob a mudança dos coeficientes a_{ij} . A observação de que a órbita mais freqüente no caso não-homogêneo ajusta-se melhor ao ciclo celular indica-nos que a escolha

homogênea talvez não seja a que aproxime melhor uma porta lógica do tipo aditiva que medeia a ação conjunta de múltiplas interações. Seria necessária uma investigação entre a relação dos coeficientes que geram essa órbita mais freqüente e como se dão as interações in vivo. Contudo isso ainda não foi realizado por nós e, mesmo nos trabalhos mais recentes, essas informações são escassas [77]. Caso esse conjunto de coeficientes seja os mais adequados, então o modelo não-homogêneo forneceria um sistema dinâmico mais robusto em relação à execução do ciclo celular: 10,9% da amostra apresenta a órbita Orb_{freq} enquanto apenas 3,6% apresenta a órbita do modelo homogêneo.

Capítulo 8

Ciclo circadiano

Observações sobre o comportamento de um inseto, a mosca *Drosophila melanogaster*, indicaram que alguns comportamentos são restritos a momentos específicos do dia. A mosca adulta ao completar seu desenvolvimento na pupa (espécie de casulo em que a mosca executa fases intermediárias de seu desenvolvimento) sai de sua casa protetora e torna-se sujeita às intempéries do mundo. Ainda sem uma formação completa de seu revestimento, especificamente sem o endurecimento da cutícula que reveste as asas, poderia sofrer um risco alto de desidratação, caso a saída da pupa ocorresse no meio do dia, quando a temperatura é a maior. A fim de evitar esse risco, as moscas adultas sempre emergem de suas pupas durante o alvorecer do dia. Mesmo se as pupas forem recolhidas em laboratório e expostas à ausência de luz, a eclosão das moscas dá-se quando elas acreditam ser o alvorecer do dia. Não só a eclosão da pupa, mas atividades executadas pela mosca adulta estão vinculadas ao ritmo dia-noite. O comportamento de alimentação, que envolve o funcionamento do sistema de sensibilidade química que localiza comida, do sistema de ativação dos músculos das asas, entre outros, também estão vinculados ao ritmo dia-noite [78].

Do latim *circa*, que significa ‘em torno’, e *dies*, que significa ‘dia’, cunhou-se o termo circadiano para os ritmos biológicos que estão vinculados ao ritmo dia-noite. Não só a mosca, mas grande parte dos organismos vivos apresentam ritmos circadianos.

A questão que surgiu dessas observações comportamentais era sobre o mecanismo endógeno de tais ritmos. Haveria um sistema de controle reponsável pela regulação dos processos biológicos de forma que esses estivessem vinculados ao dia e à noite?

Com a descoberta de técnicas de manipulação dos genes, foi possível realizar uma mutação em um gene específico e observar as mudanças fenotípicas (mudanças na aparência ou comportamento). Se um gene A é responsável por um comportamento B e uma mutação no gene A altera o comportamento B, então se pode concluir que o gene A faz parte do mecanismo genético que controla o fenótipo B. Estudando o ritmo de eclosão da pupa como sendo o comportamento B, pôde-se identificar inicialmente alguns genes responsáveis pelo controle dos ritmos circadianos. Ao primeiro gene descoberto deu-se o nome genérico ‘period’, cuja abreviação é *per*. Não só o ritmo de eclosão, mas também a atividade locomotora foi usada como o comportamento B. O mesmo gene foi identificado, o que levou à conclusão da possível existência de um único sistema de controle genético responsável pelos ritmos circadianos em geral. Posteriormente outro genes pertencentes a esse sistema de controle foram identificados [78].

8.1 Panorama geral

Basicamente, o controle da expressão gênica faz com que a produção cíclica de algumas proteínas e RNA conduza os vários processos biológicos de maneira a gerar um ritmo circadiano. Esses genes formam uma rede regulatória a que chamaremos oscilador circadiano. Esse oscilador não só coordena alguns ritmos, mas também recebe informação do ciclo dia-noite através da captação da luz. Dessa forma o oscilador vincula os ritmos fisiológicos ao ritmo externo dia-noite.

Em um organismo multicelular, várias células possuem osciladores circadianos. Em vista da coexistência de múltiplos osciladores em um mesmo organismo, algumas questões se impõem: (i) cada tipo de célula possui mecanismos genéticos distintos? (ii) Como os vários osciladores interagem entre si a fim de gerar um ritmo coordenado? (iii) Como o oscilador recebe as informações da luminosidade externa? (iv) Como a oscilação dos genes transmite a informação que garante a ritmicidade de algumas atividades biológicas?.

A resposta à primeira questão é simples: todas as células parecem utilizar o mesmo mecanismo genético. Mesmo entre espécies diferentes, há certa conservação [79], indicando que um mesmo mecanismo evoluiu preservando as principais características.

Já a segunda e a terceira questões dependem da espécie em consideração. Nos multicelulares em geral, há um conjunto de osciladores principais, que no caso dos mamíferos, se situam no núcleo supraquiasmático (conjunto de células nervosas do hipotálamo), e osciladores periféricos, que se situam em diversos tecidos. Os osciladores principais recebem a informação da luminosidade e comunicam-na aos osciladores periféricos via informações químicas. Dessa forma os osciladores principais funcionam como um marca-passo. Contudo há algumas exceções. A mosca *D. melanogaster* não apresenta um marca-passo principal, mas todas as células podem receber informação da luz diretamente [80]. O mecanismo de como a informação da luz é captada pode ser encontrada em [81, 82]. Contudo o problema de como os osciladores interagem entre si continua sem uma boa descrição [83, 84, 85, 86, 87, 88, 89, 90, 91, 92].

A resposta para a quarta questão é ainda menos conhecida. Observa-se que os genes presentes no oscilador circadiano são responsáveis por alguns ritmos circadianos, mas a especificação dos mecanismos que executam esse controle ainda não é clara [91].

Nesse capítulo somente estudaremos a questão (i). Em particular, estudaremos o oscilador circadiano da mosca *D. melanogaster*. Essa mosca é um organismo modelo no qual os esforços de pesquisa se concentram.

8.2 O oscilador circadiano da *D. melanogaster*

8.2.1 A rede regulatória

A despeito de não se ter uma compreensão dos mecanismos moleculares envolvidos na interação entre o oscilador circadiano e as atividades biológicas que apresentam ritmos circadianos, o fato é que as células possuem um oscilador central que têm compartilhados os mesmos componentes moleculares. Esse oscilador pode ser representado através de uma rede regulatória, como na figura 8.1. Essa rede foi obtida do artigo *Transcriptional Feedback Oscillators: Maybe, Maybe Not...* [93].

Nessa rede, os nós representam genes, RNA e proteínas. Embora o significado de uma ligação continue sendo o de ativação e o de inibição, a representação simultânea de genes, RNA e proteínas carecem de

no promotor desses genes; o complexo PER/TIM, em algum momento, entraria no núcleo e se ligaria ao CYC/CLK inibindo a transcrição de seus próprios genes, o *per* e o *tim*. Aliado a esse circuito de ativação e de inibição, um sistema complexo de fosforilação coordenaria os processos de entrada e saída do núcleo, de degradação e de vinculação à luminosidade externa de tal forma que o oscilador gerasse um ciclo vinculado ao ciclo dia-noite [95].

Contudo constatou-se que a concentração da CLK é relativamente constante durante o ciclo, ao passo que os RNA transcritos a partir do gene *clk* apresentam a oscilação típica do oscilador circadiano, exceto por ser fora de fase em relação à oscilação do *per* e do *tim* (essa é a razão pela qual os RNA estão representados na rede). Essa oscilação da transcrição do *clk* é devido ao fato de ele ser controlado por fatores de transcrição cuja transcrição é também ativada pelo CYC/CLK. A proteína PDP1 ϵ atua como ativador e a VRI como inibidor, sendo os seus genes o *pdp1 ϵ* e o *vri*, respectivamente [95]. Estudos envolvendo mutações no gene *clk* e modificações do estado da CLK indicaram que a oscilação da transcrição do *clk* está relacionada à amplitude da oscilação, i.e., ao valor máximo das concentrações das proteínas, cujas concentrações oscilam, e à amplitude dos ritmos comportamentais, i.e., à intensidade com que um determinado comportamento demonstra as variações provenientes do ritmo circadiano. Isso parece indicar que o papel exercido pela CLK está relacionado à manutenção de uma amplitude das concentrações das demais proteínas, forte o suficiente para garantir a funcionalidade do oscilador. Além do mais, parece que a oscilação da transcrição do *clk* garante uma estabilidade ao nível de concentração da CLK, haja visto os eventos de fosforilação que exercem um forte controle nesse nível [95].

Em suma, podemos dizer que há dois circuitos interconectados: (i) o primeiro composto pelos genes *per*, *tim* e pelas proteínas PER, TIM, CYC e CLK e (ii) o segundo composto pelos genes *vri*, *pdp1 ϵ* , *clk* e pelas proteínas VRI, PDP1 ϵ , CLK e CYC. O primeiro circuito determina a oscilação propriamente dita, enquanto o segundo garante a funcionalidade do oscilador, ou mesmo uma estabilidade [93]. Ambos circuitos se entrelaçam através da formação do complexo CYC/CLK/PER/TIM, quando se dá a inibição do CYC/CLK pelo PER/TIM. Apesar de o controle pós-transcricional ser de extrema relevância para o funcionamento do oscilador, em vista da precariedade de informações sobre os eventos de fosforilação, buscamos uma representação apenas do sistema de controle transcricional, que está representado na rede regulatória na figura 8.1. As oscilações das concentrações de algumas moléculas *in vivo* durante uma oscilação completa correspondente a 24hs estão na figura 8.2.

8.2.3 O modelo

Da mesma forma que no estudo do ciclo celular, vamos vincular à rede o mesmo sistema dinâmico discutido anteriormente. Mas, no presente caso, além da inativação implícita, vamos acrescentar uma ativação implícita a todos os nós que não possuem nenhuma ligação de ativação: se $\sum_j a_{ij}\sigma_j(t) = 0$ durante o intervalo $t_0 \leq t \leq t_0 + t_d$ em que o nó i está desativado ($\sigma_i(t) = 0$), então a proteína será ativada no tempo $t_0 + t_d$ ($\sigma_i(t_0 + t_d)$). Durante todo o restante do trabalho, assumiremos $t_d = 0$.

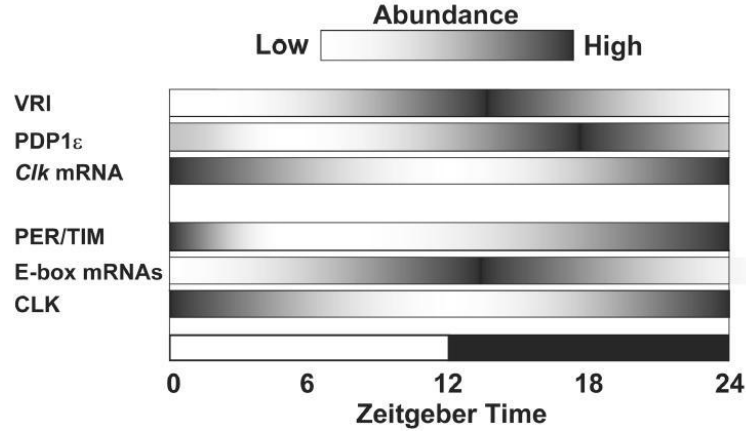


Figura 8.2: Concentrações de algumas proteínas e RNA em função do tempo para um período do ciclo circadiano [6].

As regras da evolução temporal do sistema dinâmico, escritas explicitamente, são

$$\begin{aligned}
 \sigma_1(t+1) &= F(\sigma_{11}(t) - \sigma_5(t)) & \sigma_2(t+1) &= F(\sigma_1(t)) \\
 \sigma_3(t+1) &= F(\sigma_1(t)) & \sigma_4(t+1) &= F(\sigma_1(t)) \\
 \sigma_5(t+1) &= F(\sigma_8(t) + \sigma_9(t)) & \sigma_6(t+1) &= F(\sigma_{10}(t)) \\
 \sigma_7(t+1) &= F(-\sigma_6(t)) & \sigma_8(t+1) &= F(\sigma_2(t)) \\
 \sigma_9(t+1) &= F(\sigma_3(t)) & \sigma_{10}(t+1) &= F(\sigma_4(t)) \\
 \sigma_{11}(t+1) &= F(\sigma_7(t)) & &
 \end{aligned} \tag{8.1}$$

A rede apresenta nós de vários tipos: genes, RNA e proteínas. Os significados dos termos ativado e inativado precisam ser esclarecidos. Uma proteína ativada é realmente uma proteína que está numa conformação ativa, mas um RNA ativado significa somente um nível de RNA suficiente e em condições para que haja a tradução da proteína. O gene ativado evidentemente significa um gene cujo fator de transcrição está ligado às regiões de ativação desse gene.

A degradação implícita se dá por motivos de ausência de alguns circuitos de inibição, mas também por consistência de definição. As degradações da TIM e PER se dão através de um complexo de degradação que reconhece proteínas devidamente marcadas, mas como esses complexos estão ausentes na rede utilizada, é necessário que o nó PER/TIM tenha uma inativação implícita. Por um motivo similar, é necessário impor inativação implícita na VRI. A inativação implícita dos demais nós se dá por motivos de consistência. Se o fator de transcrição CYC/CLK está inativo, então os genes *per*, *tim* e *vri* não serão transcritos. Mas suponha-se que no tempo t o *per* esteja ativo ($\sigma_2(t) = 1$) e no tempo $t+1$ o complexo CYC/CLK seja inibido ($\sigma_1(t+1) = 0$). Na ausência de inativação implícita, teríamos $\sigma_2(t+2) = 1$, o que é inconsistente com o fato de não mais haver o fator de transcrição. Nesse caso, é necessário impor a inativação implícita. O caso dos nós correspondente aos RNA é similar. Só há RNA se o gene estiver sendo transcrito, i.e., ativo.

Já a ativação implícita é imposta a todos os nós que não possuem nenhuma ligação de ativação. Note-se que no caso da rede regulatória do ciclo celular discutido no capítulo anterior, o nó correspondente

Tempo	CYC-CLK	<i>per</i>	<i>tim</i>	<i>vri</i>	PER/TIM	VRI	<i>clk</i>	<i>vri</i> RNA	<i>per</i> RNA	<i>tim</i> RNA	<i>clk</i> RNA
1	0	0	0	0	0	0	1	0	0	0	0
2	0	0	0	0	0	0	1	0	0	0	1
3	1	0	0	0	0	0	1	0	0	0	1
4	1	1	1	1	0	0	1	0	0	0	1
5	1	1	1	1	0	0	1	1	1	1	1
6	1	1	1	1	1	1	1	1	1	1	1
7	1	1	1	1	1	1	0	1	1	1	1
8	1	1	1	1	1	1	0	1	1	1	0
9	0	1	1	1	1	1	0	1	1	1	0
10	0	0	0	0	1	1	0	1	1	1	0
11	0	0	0	0	1	1	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0

Tabela 8.1: Órbita correspondente ao ciclo de período 12 que é o único atrator da dinâmica.

ao *Cln3* não possui nenhuma ativação por parte de proteínas pertencentes à rede. Mas, nesse caso, não teria significado biológico impor ativação implícita, pois o que ocorre não é um mecanismo ausente, mas simplesmente o sinal de ativação é externo à rede (tamanho da célula). No presente caso, a situação é mais sutil. O único nó que necessita de ativação implícita é o gene *clk*. É sabido que o nível de VRI aumenta juntamente com o aumento de seu RNA, mas o nível do PDP1 ϵ sofre um atraso em relação ao nível do seu RNA. Dessa forma, se os RNA do *vri* e do *pdp1 ϵ* forem transcritos, isso não significa que as proteínas VRI e PDP1 ϵ estarão imediatamente ativas e em nível ideal de atividade. Além disso, mesmo se pequenas quantidades de PDP1 ϵ estiverem presentes, na ausência de VRI, a PDP1 ϵ irá ativar o *clk* [79]. A rede não captura essas características relevantes. Se o nó correspondente ao PDP1 ϵ for inserido na rede juntamente com seu gene e RNA e as regras gerais de atribuição de inativação e ativação implícitas forem usadas, obtém-se um único ponto fixo na dinâmica. Mas, se nós excluirmos a interação envolvendo esses nós, teremos que impor uma ativação implícita no *clk*. Essa ativação implícita incorpora o fato de que o *clk* estará sempre ativo na ausência do VRI, quando então o PDP1 ϵ , mesmo em quantidades pequenas, irá fazer o trabalho de ativação. Além disso, se incluirmos na rede o circuito Pdp1 ϵ e mantivermos a ativação implícita do Clk, o ponto fixo desaparece e teremos 1 ciclo. Portanto a ativação implícita está mais de acordo com o processo biológico *in vivo*.

Tendo feito essas considerações iniciais, procedamos à análise do sistema dinâmico e suas propriedades.

8.3 A estabilidade e a robustez do oscilador circadiano da *D. melanogaster*

Com a utilização de um simples programa computacional, constatamos que o sistema dinâmico possui um único atrator: um ciclo de período 12, que está na tabela 8.1.

Pode-se ver que o RNA do *clk* oscila fora de fase com os RNA do *per* e *tim*, em acordo com as oscilações presentes na figura 8.2, apesar de detalhes mais finos não terem sido evidenciados. Evidentemente essa discrepância se deve ao caráter simplificado do modelo. Quando o estado do nó PER/TIM é igual a 1, não significa somente que o PER/TIM foi produzido, mas também significa que ele já se encontra no núcleo, apto a se ligar ao CYC/CLK. O fluxo de entrada é controlado por eventos de fosforilação que não foram incorporados no modelo [79], por isso o ciclo apresenta ambos VRI e PER/TIM oscilando na mesma fase, embora *in vivo* isso não ocorra. O nó PER/TIM ativado significa que ambas proteínas já foram produzidas, já foram translocadas para o núcleo e já se encontram em estados de fosforilação ativos [94]. Esse processo implicaria o atraso em relação à oscilação da VRI. Além do mais, como o

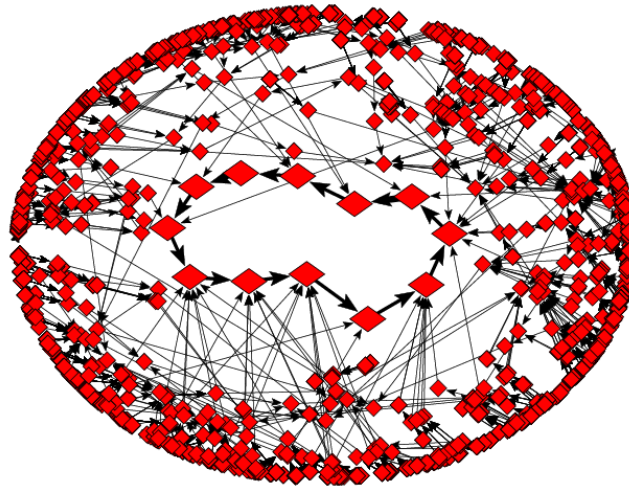


Figura 8.3: Órbitas no espaço de estados. O ciclo, que é o único atrator, está representado no centro da figura com os nós e setas maiores.

modelo só assume estados discretos, não é possível fazer uma comparação detalhada com os níveis de concentração, pois os estados indicam apenas concentração acima ou abaixo da qual o nó está ativado ou inativado. Apesar dessas limitações, temos um ciclo biologicamente significativo. Vejamos as informações que podemos extrair do modelo.

O resultado principal é que todas as 2048 configurações iniciais são atraídas para o mesmo e único ciclo, o da tabela 8.1. Na figura 8.3 mostramos o espaço de estados com as órbitas sendo atraídas para o ciclo de período 12.

A fim de verificar se a presença de uma bacia de atração grande é fruto da arquitetura particular da rede regulatória ou se é uma característica compartilhada por redes aleatórias, pode-se construir redes aleatórias equivalentes e calcular o tamanho das bacias de atração dos pontos fixos. No presente caso, a rede aleatória deve preservar as principais características da rede real: (i) mesmo número de nós e ligações, (ii) cada nó deve ter ao menos uma ligação incidente de maneira a garantir que todos os nós estejam participando da dinâmica, (iii) não deve haver mais que uma ligação entre dois nós e (iv) a rede não deve apresentar um conjunto de nós desconectado dos demais (a rede deve ser conexa). Utilizando um ensemble suficientemente grande, constatamos que apenas 5,7% do ensemble apresenta um período maior do que 8 e com bacia de atração maior do que 1434 (70% das configurações iniciais). Isso indica que a propriedade de todo o espaço de estados ser atraído para um único atrator não é um propriedade compartilhada pelas redes aleatórias.

De acordo com o conceito de estabilidade introduzido na seção 6.1, o ponto fixo é extremamente estável, pois todas as 2048 configurações iniciais convergem para o ciclo que corresponde ao ciclo do oscilador circadiano. Não importa que o oscilador tenha sofrido uma perturbação, ele sempre volta a oscilar no mesmo ciclo sem o risco de ficar preso a um outro atrator, que pode representar uma oscilação possivelmente perigosa para a célula. Essa estabilidade é realmente um atributo da rede do oscilador circadiano e não um atributo geral das redes aleatórias.

Uma característica interessante é a presença dos dois circuitos interconectados. Esses circuitos são conhecidos como “feedbacks” negativos, pois eles são do tipo: A ativa B, B ativa C e C inibe A. O resultado

posterior de um evento regula o mesmo. Podemos perguntar se esse par de circuitos é necessário para garantir a estabilidade. A fim de manter a clareza do termo estabilidade, lembramos que estabilidade significa, em nosso caso, simplesmente uma bacia de atração grande o suficiente para que trajetórias perturbadas tenham grande chance de permanecer na mesma bacia de atração. Já foi observado que o segundo circuito não é tão imprescindível para que haja oscilação [93, 95]. Num experimento transgênico [96], o *clk* é controlado pelos mesmos elementos regulatórios do *per*. Nesse caso, nenhum efeito catastrófico foi constatado, observando-se apenas algumas alterações pequenas no período e fase dos constituintes. A fim de modelar esse experimento transgênico, uma representação alternativa da rede pode ser feita: excluímos o segundo circuito, adicionamos ativações do CYC/CLK para o *clk* e mantivemos as mesmas degradações e ativações implícitas como se a rede fosse a original.

Nesse caso o espaço de estados possui $2^8 = 256$ estados e a dinâmica apresenta um único ponto fixo. Nenhum ciclo aparece. Mas se modificarmos ligeiramente o valor do coeficiente a_{ij} correspondente à ligação PER/TIM \rightarrow CYC/CLK, colocando $a_{1,6} = -2$, então surgem 2 ciclos de período 8. Isso introduz a questão do papel do segundo circuito, haja vista que em sua ausência a estabilidade e a robustez são diminuídas drasticamente: o sistema dinâmico sofre alteração drástica sob mudança do parâmetro a_{ij} e as bacias de atração são reduzidas. Podemos então perguntar se a presença de dois circuitos do tipo “feedback” negativo realmente aumenta a estabilidade e, evidentemente, a complexidade de uma maneira geral ou se é apenas uma característica da rede da *Drosophila*. Para responder essa questão nós construímos três tipos de ensembles de redes aleatórias: com uma, duas ou nenhuma ligação de inibição, sendo as demais ligações de ativação. Com essa construção podemos obter redes com circuitos do tipo “feedback” negativo interconectados. Nesses ensembles surgem pontos fixos e ciclos, mas somente os ciclo com período próximo ao do oscilador circadiano e com bacia de atração grande seriam bons candidatos a executar uma performance de um oscilador circadiano. De acordo com o que mencionamos anteriormente, a probabilidade de termos um ciclo razoável como resultado de eventos aleatórios é de apenas 5,7%. Mas nós podemos perguntar se as redes com quantidades crescentes de circuitos do tipo “feedback” negativo aumentam essa probabilidade, i.e., se aumentar a complexidade pode fornecer maior estabilidade para a performance da dinâmica. Procedendo à análise dos três ensembles, não importando a quantidade de inibições presentes, todas as redes apresentam em torno de 4 ciclos distintos (considerando o ponto fixo como um ciclo de período 1). Mas a probabilidade de ser ter um ciclo com período maior que 8 e bacia de atração maior que 1434 (70% de 2048) é 0%, 3% e 5,7% para nenhuma, uma ou duas inibições, respectivamente. Por outro lado, a probabilidade de se ter um ponto fixo como bacia de atração maior que 1434 é 43%, 27% e 19%. Esses resultados sugerem que a presença de um ciclo de período razoavelmente grande e estável é uma propriedade atingida pelo aumento da complexidade através da presença de circuitos “feedback” negativos. Dessa forma, o segundo circuito no oscilador da *Drosophila* parece estar presente para garantir um bacia de atração grande e um período relativamente grande, i.e., para garantir estabilidade.

Agora vamos discutir a robustez da rede em relação a alteração dos coeficientes a_{ij} . A única maneira não redundante de inserir a não-homogeneidade é alterando o valor dos coeficientes relacionados à dinâmica do nó CYC/CLK, i.e., colocando (i) $a_{ij} = -2$, em que i =CYC/CLK e j =PER/TIM, ou colocando (ii) $a_{ij} = +2$, em que i =CYC/CLK e j =PER/TIM. No caso (i) um único ciclo de período 10 emerge e no caso (ii) um ciclo de período 12, similar ao do caso homogêneo, emerge juntamente com um ciclo de período 5 com bacia de atração pequena. Isso indica que a rede mantém seu desempenho de um oscilador com um

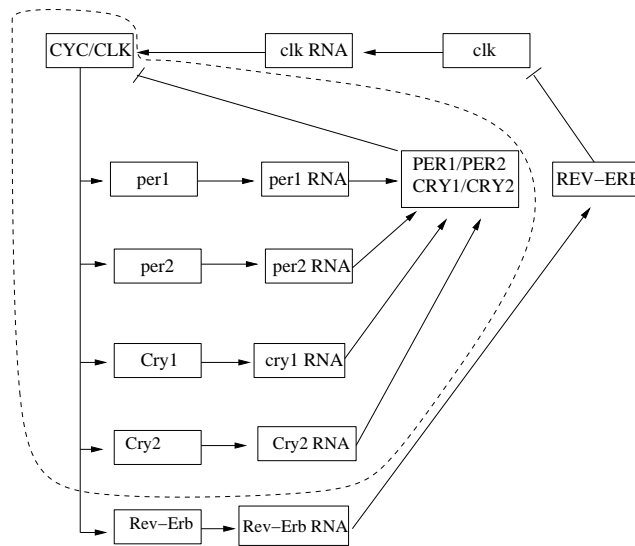


Figura 8.4: Rede regulatória do oscilador circadiano dos mamíferos. A linha tracejada separa os dois osciladores entrelaçados.

ciclo de bacia de atração grande, podendo ser, nesse sentido, considerada robusta.

Vejam os a robustez em relação à eliminação de algumas ligações. O sistema dinâmico não sofre alteração se deletarmos qualquer ligação de ativação do primeiro circuito, pois há duas formas independentes de ativar o PER/TIM. Mas note-se que, do ponto de vista biológico, a proteína PER é essencial para a manutenção do ciclo, e que a TIM é necessária para garantir maior estabilidade ao PER [95]. Se uma das ligações de ativação do PER/TIM for eliminada, o nosso sistema dinâmico continua a reproduzir a oscilação circadiana, o que não é verdade *in vivo*. Essa é uma robustez do sistema dinâmico que não possui correspondência com uma robustez biológica, o que é devido ao fato de o sistema dinâmico admitir uma porta lógica do tipo aditiva. Por outro lado, não há ciclos se alguma ligação do segundo circuito for eliminada. Note-se que esse segundo circuito não possui uma parte redundante e, por isso, não é robusto nem do ponto de vista formal do sistema dinâmico. Talvez a robustez do oscilador possa ser obtida através do acoplamento entre vários osciladores [87].

8.4 O ciclo circadiano dos mamíferos

Muitos dos genes que controlam os ritmos circadianos em diferentes espécies se mantiveram correlacionados através do tempo (são genes homólogos). Essa correlação garantiu a conservação do mecanismo básico desses osciladores, como pode ser visto através da comparação entre a rede da drosófila e dos mamíferos 8.4, em que a espécie modelo utilizada é o rato.

Os significados biológicos das interações são análogos aos da drosófila, portanto nos absteremos de maiores comentários [93, 79, 82].

O estudo do sistema dinâmico mostrou que todas as $2^{15} = 32768$ configurações iniciais são atraídas para um único atrator. Esse atrator é um ciclo de período 12 similar ao atrator da drosófila. As mesmas análises realizadas no caso da drosófila forneceram resultados qualitativamente análogos para o caso dos mamíferos. Em particular, no ensemble de redes aleatórias equivalentes, apenas 5,4% de redes possuem

ciclo com período maior que 8 e bacia de atração maior que 70% do espaço de estados.

É interessante que, novamente, o comportamento temporal de uma rede regulatória, bem como o estudo de estabilidade e robustez possam ser capturados por um sistema dinâmico tão simples. Apesar de as simplificações desse modelo conduzirem a algumas discrepâncias em relação aos dados biológicos, características gerais podem ser bem trabalhadas nesse ambiente formal.

Capítulo 9

Conclusão

Apesar de os sistemas biológicos se organizarem de maneira complexa, em que vários elementos interagem entre si a fim de executar uma determinada tarefa, algumas propriedades podem ser estudadas de maneira simples. Com a representação do sistema de controle transcricional e pós-transcricional em termos das redes regulatórias e com o modelamento da evolução temporal através de um sistema dinâmico discreto simples, fomos capazes de estudar propriedades relevantes dos sistemas biológicos: a robustez e a estabilidade.

O ciclo celular demonstrou ser robusto em relação à modificação dos parâmetros $a_{i,j}$, que determinam o peso das ligações, bem como em relação ao parâmetro t_d que determina o tempo de degradação. O atrator da dinâmica, que representa a seqüência de estados por que passa a célula durante o ciclo celular, no caso do modelo homogêneo, demonstrou ser estável, no sentido de que atrai 86% das configurações iniciais. Contudo, no modelo em que as não-homogeneidades são incorporadas, surge um atrator mais estável, que atrai 91% das configurações iniciais. Não só possui uma maior estabilidade, mas também apresenta um atrator cujo significado biológico é mais coerente com os sistemas *in vivo*. O interessante é notar que um modelo simples foi capaz de caracterizar o ciclo celular como sendo robusta e estavelmente construído.

Similarmente, o oscilador circadiano mostrou ser um oscilador extremamente estável, atraindo 100% das configurações iniciais para o ciclo que caracteriza o oscilador. A presença de dois circuitos do tipo “feedback” negativo interconectados parece estar relacionado à estabilidade do oscilador e não com a capacidade de oscilar. Na ausência de um dos circuitos, apesar de a capacidade oscilatória ser mantida, surgem dois ciclos, o que diminui a estabilidade. Não só a rede do oscilador circadiano, mas também redes que em geral possuem dois circuitos “feedback” negativos apresentam ciclos de período grande com bacias de atração também grandes, em contraposição aos que possuem somente um ou nenhum. Ao contrário do ciclo celular, o oscilador circadiano parece não ser muito robusto. A evolução parece ter conservado o esquema básico de regulação das oscilações circadianas, de maneira tal que tanto o oscilador da drosófila quanto o dos mamíferos apresentam propriedades qualitativamente similares.

A teoria de redes complexas fornece um sistema formal em que a estrutura e a dinâmica das redes podem ser utilizadas no estudo das redes complexas em biologia.

Referências Bibliográficas

- [1] Réka Albert. Scale-free networks in cell biology. *Journal of Cell Science*, 118:4947–4957, 2005.
- [2] H. Ma and Zeng A.-P. Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. *Bioinformatics*, 2003.
- [3] Tong Ihn Lee et al. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science*, 298:799–804, 2002.
- [4] D.J. Watts and S.H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.
- [5] Fangtin Li, Tao Long, Ying Lu, Qi Ouyang, and Chao Tang. The yeast cell-cycle network is robustly designed. *PNAS*, 101:4781–4786, 2004.
- [6] Paul E. Hardin. Transcription regulation within the circadian clock: The e-box and beyond. *Journal of Biological Rhythms*, 19:348–360, 2004.
- [7] E.Klipp, R. Herwig adn A. Kowald, C. Wierling, and H. Lehrach. *Systems Biology in Practice*. Wiley-VCH, 2005.
- [8] David Eisenberg, Edward M. Marcotte, Ioannis Xenarios, and Todd O. Yeates. Protein function in the post-genomic era. *Nature*, 405:823–826, 2000.
- [9] L.H. Hartwell, J.J. Hopfield, S. Leibler, and A.W. Murray. From molecular to modular cell biology. *Nature*, 402:C47–C52, 1999.
- [10] Zhi Wang and Jianzhi Zhang. In seach of the biological significance of modular structure of protein networks. *PLoS Computational Biology*, 3:1011–1021, 2007.
- [11] Roger Guimera and L.A.N. Amaral. Functional cartography of complex metabolic networks. *Nature*, 433:895–900, 2005.
- [12] Roded Sharan, Igor Ulitsky, and Ron Shamir. Network-based prediction of protein function. *Molecular System Biology*, doi:10.1038/msb4100129, 2007.
- [13] Pavlos Alifragas, Gretchen Pootinga, Susan M. Parkhurst, and Christos Delidakis. A network of interacting transcriptional regulators involved in *Drosophila* neural fate specification revealed by the yeast two-hybrid system. *PNAS*, 94:13099–13104, 1997.

-
- [14] H.Jeong, B.Tombor, R. Albert, Z.N. Oltvai, and A.-L. Barabási. The large-scale organization of metabolic networks. *Nature*, 407:651–654, 2000.
- [15] A.-L. Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.
- [16] Panos Oikonomou and Philippe Cluzel. Effects of topology on network evolution. *Nature Physics*, 2:532–536, 2006.
- [17] Evan A. Variano, Jonathan H. McCoy, and Hod Lipson. Networks, dynamics, and modularity. *Physical Review Letters*, 92:18701, 2004.
- [18] Stephen Jay Gould. *Wonderful Life: The Burgess Shale and the Nature of History*. W.W.Norton& Co., 1989.
- [19] Béla Bollobás. *Graph Theory, an Introductory Course*. Springer-Verlag, 1985.
- [20] J.F.F. Mendes and S.N. Dorogovtsev. *Evolution of Networks, From Biological Nets to the Internet and WWW*. Oxford University Press, 2003.
- [21] Radu Balescu. *Equilibrium and Nonequilibrium Statistical Mechanics*. John Wiley Sons, 1975.
- [22] Sílvio R. A. Salinas. *Introdução à Física Estatística*. Editora da Universidade de São Paulo, 2005.
- [23] R. Albert, H.Jeong, and A.-L. Barabási. Diameter of the world wide web. *Nature*, 401:130–131, 1999.
- [24] R. Albert, A.-L. Barabási, H. Jeong, and G. Bianconi. Power-law distribution of the world wide web. *Science*, 287:2115a, 2000.
- [25] A.-L. Barabási. The physics of the web. *Physics World*, July 2001.
- [26] Z. N. Oltvai and A.-L. Barabási. Life’s complexity pyramid. *Science*, 298:763–764, 2002.
- [27] S.-H. Yook, H. Jeong, and A.-L. Barabási. Modeling the internet’s large-scale topology. *PNAS*, 99:13382–13386, 2002.
- [28] A. Broader et al. Graph structure in the web. *Comput. Netw.*, 33:309, 2000.
- [29] M.E.J. Newman. Small worlds: The structure of scientific colaboration networks. *PNAS*, 98:404, 2001.
- [30] R. Kumar et al. The web as a graph. *Procedings fo the 19th ACM Symposium on Priniples of Database Sytems*, pages 1–10, 2000.
- [31] D. M. Penock et al. Winners don’t take all: Characterising the competition for links on the web. *PNAS*, 99:5207, 2002.
- [32] S. Redner. How popular is your paper? *Eur. Phys. J. B*, 23:267, 1998.
- [33] L. Barabási A. Evolution of the social network os scientific collaborations. *Physica A*, 311:590, 2002.
- [34] F. Lijjeros et al. The web of human sexual contacts. *Nature*, 411:907–, 2001.

-
- [35] J.M. Montoya and R.V. Solé. Small world patterns in food webs. *J. Teor. Biol.*, 214:405, 2002.
- [36] W. Aiello and F. Chung and L. Lu. A random graph model for massive graphs. *Proceedings of the Thity-second Annual ACM Symposium on Theory of Computing*, pages 171–180, 2000.
- [37] L.-I. Mielsch H. Ebel and S Bornholdt. Scale-free topology of e-mail networks. *Phys. Rev. E*, 66, 2002.
- [38] R. Albert, H. Jeong, and A.-L. Barabási. Error and attack tolerance in complex networks. *Nature*, 406:378–381, 2000.
- [39] L.A. Barbosa, A. Castro e Silva, and J. Kamphorst Leal da Silva. Scalling relations in food webs. *Physical Review E*, 73:041903, 2006.
- [40] L. Giot et al. A protein interaction map of drosophila melanogaster. *Science*, 302:1727–1736, 2003.
- [41] P. Uetz et al. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature*, 403:623–627, 2000.
- [42] E. Almaas, B. Kovacs, T. Vicsek, Z.N. Oltvai, and A.-L. Barabási. Global organization of metabolic fluxes in the bacterium escherichia coli. *Nature*, 427:839–843, 2004.
- [43] Z.N. Oltvai E. Almaas and A.-L. Barabási. The activity reaction core and plasticity of metabolic networks. *PLoS Computational Biology*, 1:0557–0563, 2005.
- [44] I. Farkas, H. Jeong, T. Vicsek, A. L. Barabási, and Z. N. Oltvai. The topology of the transcription regulatory network in the yeast, *Saccharomuces cerevisiae*. *Physica A*, 318:601–612, 2003.
- [45] Piotr Bialas, Zdzislaw Burda, and Des Johnston. Condensation in the backgammon model. *Nuclear Physics B*, 493.
- [46] S. N. Dorogovtsev, J.F.F Mendes, A. M. Povolosky, and A. N. Samukhin. Organization of complex networks without multiple connections. *Physical Review Letters*, 95, 2005.
- [47] M.E.J. Newman and D.J.Watts. Renormalization group analysis of the small-world network model. *Physcis Letters A*, 263:341–346, 1999.
- [48] Rueven Cohen and Shlomo Havlin. Scale-free networks are ultrasmall. *Physical Review Letters*, 90:058701, 2003.
- [49] A.-L. Barabási and Z.N. Oltvai. Network biology: Understanding the cell’s functional organization. *Nature Reviews*, 5:101–113, 2004.
- [50] P.L.Krapivsky and S.Redner. Organization of growing networks. *Physical Review E*, 63:066123, 2001.
- [51] S.-H. Yook, Z.N.Oltvai, and A.-L. Barabási. Functional and topological characterization of protein interaction networks. *Proteomics*, 4:928–942, 2004.
- [52] E. Ravasz and A.-L. Barabási. Hierarchical organization in complex network. *Physical Review E*, 67:026112, 2003.

-
- [53] Jing-Dong J. Han et al. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature*, 430:88–93, 2004.
- [54] E.Ravasz, A.L. Somera, D.A. Mongru, Z.N. Oltvais, and A.-L. Barabási. Hierarchical organization of modularity in metabolic networks. *Science*, 297:1551–1555, 2002.
- [55] Claus I. Doering and Artur O. Lopes. *Equações Diferenciais Ordinárias*. IMPA, 2005.
- [56] John J. tyson, Kathy Chen, and Bela Novak. Network dynamics and cell physiology. *Nature Reviews*, 2:908–916, 2001.
- [57] Bruce Alberts. *Biologia Molecular da Célula*. Artmed, 2002.
- [58] Jacob H. van L. *Coding Theory*. Springer-Verlag, 1973.
- [59] Uri Alon. Network motifs: Theory and experimental approaches. *Nature Reviews*, 450:450–461, 2007.
- [60] Uri Alon. *An Introduction to System Biology*. Chapman & Hall/CRC, 2007.
- [61] Towards a theory of biological robustness. The yeast cell-cycle network is robustly designed. *Molecular System Biology*, Editorial, 2007.
- [62] Alexander W. Rives and Timothy Galitski. Modular organization of cellular networks. *PNAS*, 100:1128–1133, 2003.
- [63] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69:026113, 2004.
- [64] M.E.J. Newman. Detecting community structure in networks. *The European Physical Journal B*, 38:321–330, 2004.
- [65] Rosela Vinsintin et al. Cdc20 and cdh1: a family of substrate-specific activators of apc-dependent proteolysis. *Science*, 278:460–463, 1997.
- [66] Michael Schwab, Annegret Schulze Lutum, and Wolfgang Seufert. Yeast hct1 is a regulator of clb2 cyclin proteolysis. *Cell*, 90:683–693, 1997.
- [67] Foong May Yeong, Hong Hwa Lim, Padmashree C. G., , and Uttam Surana. Exit from mitosis in budding yeast: Biphasic inactivation of the cdc28-clb2 mitotic kinase and the role of cdc20. *Molecular Cell*, 5:501–511, 2000.
- [68] Jürg Bähler. Cell-cycle control of gene expression in budding and fission yeast. *Annu. Rev. Genet.*, 39:69–94, 2005.
- [69] Kurt W. Kohn. Molecular interaction map of the mammalian cell cycle control and dna repair systems. *Molecular Biology of the Cell*, 10:2703–2734, 1999.
- [70] Itamar Simon et al. Serial regulation of transcriptional regulators in the yeast cell cycle. *Cell*, 106:697–708, 2001.

- [71] Mauren Maher, Feng Cong, David Kindelberger, Kim Nasmyth, and Stephen Dalton. Cell cycle-regulated transcription of the *clb2* gene is dependent on *mcm1* and a ternary complex factor. *Molecular and Cellular Biology*, 15:3129–3137, 1995.
- [72] Steven I. Reed. Ratchets and clocks: the cell cycle, ubiquitylation and protein turnover. *Nature Reviews*, 4:855–864, 2003.
- [73] Mike Tyers. Cell cycle goes global. *Current Opinion in Cell Biology*, 16:602–613, 2004.
- [74] Attila Csikász-Nágyi, Dojrsuren Battogtokh, Katherine Z. Chen, Béla Novák, and John J. Tyson. Analysis of a generic model of eukaryotic cell-cycle regulation. *Biophysical Journal*, 90:4361–4379, 2006.
- [75] Foong May Yeong, Hong Hwa Lim, Padmashree C. G., and Uttam Surana. Exit from mitosis in budding yeast: Inactivation of the *cdc28/clb2* mitotic kinase and the role of *cdc20*. *Biophysical Journal*, 5:501–511, 2000.
- [76] Ralph Wäsch and Frederick R. Cross. Apc-dependent proteolysis of the mitotic cyclin *clb1,2* is essential for mitosis exit. *Nature*, 418:556–562, 2002.
- [77] Katherine C. Chen et al. Kinetic analysis of a molecular model of the budding yeast cell cycle. *Molecular Biology of the Cell*, 11:369–391, 2000.
- [78] Satchidananda Panda, John B. Hogenesch, and Steve A. Kay. Circadian rhythms from flies to human. *Nature*, 417:329–335, 2002.
- [79] W. Yu and P. E. Hardin. Circadian oscillators of drosophila and mammals. *Journal of Cell Science*, 119:4793–4795, 2006.
- [80] Deborah Bell-Pedersen et al. Circadian rhythms from multiple oscillators: Lessons from diverse organisms. *Nature Reviews*, AOP:doi:10.1038/nrd1633, 2005.
- [81] Paul E. Hardin. The circadian timekeeping system of drosophila. *Current Biology*, 15:R714–R722, 2005.
- [82] Steven M. Reppert and David R. Weaver. Coordination of circadian timing in mammals. *Nature*, 418:935–941, 2002.
- [83] M. Amdaoud, M. Vallade, C. Weiss-Schaber, and I. Mihalcescu. Cyanobacterial clock, a stable phase oscillator with negligible intercellular coupling. *PNAS*, 104:7051–7056, 2007.
- [84] María de la Paz Fernández, Jessie Chu, Adriana Vellella, Nigel Atkinson, Steve A. Kay, and María Fernanda Ceriani. Impaired clock output by altered connectivity in the circadian network. *PNAS*, 104:5650–5655, 2007.
- [85] Erik D. Herzog, Sara J. Aton, Rika Numano, Yoshiyuki Sakaki, and Hajime Tei. Temporal precision in the mammalian circadian system: a reliable clock from less reliable neurons. *Journal of Biological Rhythms*, 19:35–46, 2004.

-
- [86] Premananda Indic, William J. Schwartz, Erik D. Herzog, Nicholas C. Foley, and Michael C. Antle. Modeling the behavior of coupled cellular circadian oscillators in the suprachiasmatic nucleus. *Journal of Biological Rhythms*, 22:211–219, 2007.
- [87] Andrew C. Liu et al. Intercellular coupling confers robustness against mutations in the scn circadian clock network. *Cell*, 129:605–616, 2007.
- [88] Michael N. Nitabach. Clock coordination. *Nature*, 438:173–175, 2005.
- [89] Dan Stoleru, Ying Peng, José Agosto, and Michael Rosbash. Coupled oscillators control morning and evening locomotor behaviour of drosophila. *Nature*, 431:862–868, 2004.
- [90] Dan Stoleru¹, Ying Peng¹, Pipat Nawathea¹, and Michael Rosbash. A resetting signal between drosophila pacemakers synchronizes morning and evening activity. *Nature*, 438:238–242, 2005.
- [91] Markus Stratmann and Ueli Schibler. Properties, entrainment, and physiological functions of mammalian peripheral oscillators. *Journal of Biological Rhythms*, 21:494–506, 2006.
- [92] Tsz-Leung To, Michael A. Henson, Erik D. Herzog, and Francis J. Doyle III. A molecular model for intercellular synchronization in the mammalian circadian clock. *Biophysical Journal*, 92:3792–3803, 2007.
- [93] Patricia L. Lakin-Thomas. Transcriptional feedback oscillators: Maybe, maybe not... *Journal of Biological Rhythms*, 21:83–92, 2006.
- [94] Pablo Meyer et al. Per-tim interactions in living drosophila cells: an interval timer for the circadian clock. *Science*, 331:226–229, 2006.
- [95] Kiho Bae and Isaac Edery. Regulating a circadian clock’s period, phase and amplitude by phosphorylation: Insights from drosophila. *JB Review*, 140:609–617, 2006.
- [96] Eun Young Kim et al. Drosophila clock protein is under posttranscriptional control and influences light-induced activity. *Neuron*, 34:69–81, 2002.