

HERMAN FIALHO FUMIÃ

UM MODELO PARA REGULAÇÃO GÊNICA E A
DIFERENCIAÇÃO CELULAR NOS ESTÁGIOS INICIAIS
DA EMBRIOGÊNESE

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

VIÇOSA
MINAS GERAIS - BRASIL
2010

HERMAN FIALHO FUMIÃ

**UM MODELO PARA REGULAÇÃO GÊNICA E A DIFERENCIAÇÃO
CELULAR NOS ESTÁGIOS INICIAIS DA EMBRIOGÊNESE**

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

APROVADA: 19 de fevereiro de 2010.

Prof. Hallan Souza e Silva

Prof. Sidney Geraldo Alves

Prof. Márcio Santos Rocha

Prof. Everaldo Arashiro

Prof. Marcelo Lobato Martins
(Orientador)

Aos meus pais Ronei e Joana e à minha irmã Agatha.

“*Quid est veritas?*” Pôncio Pilatos a Cristo.

Agradecimentos

- Ao professor Marcelo e ao professor Hallan, pela orientação, paciência e amizade.
- Ao professor Silvio, por ter me apresentado a magnífica área da Física Biológica.
- Aos meus amigos de fé: Fabio, Saulo, Leonardo, Gabriel, Glaubert, Davidson e Leozinho.
- À todos os professores, por me abrirem as portas do conhecimento.
- À G. Puccini e C. Debussy, cujas obras magistrais tornam meu dia menos enfadonho.
- À CAPES, FAPEMIG e ao REUNI, pelo apoio financeiro.

Sumário

| | |
|---|-------------|
| Sumário | v |
| Resumo | vii |
| Abstract | viii |
| 1 Introdução | 1 |
| 2 Uma Breve Revisão Biológica | 3 |
| 2.1 DNA: a Molécula da Vida | 3 |
| 2.2 A Estrutura do DNA | 4 |
| 2.3 Transcrição e Tradução: a Execução do Código | 5 |
| 2.4 As Informações Contidas na Rede de Regulação Gênica | 7 |
| 2.5 A Expressão Gênica e os Diferentes Tipos Celulares | 8 |
| 2.6 A Biologia do Desenvolvimento | 9 |
| 2.7 Etapas do Desenvolvimento Embrionário | 10 |
| 2.7.1 Clivagem | 10 |
| 2.7.2 Gastrulação | 11 |
| 2.7.3 Neurulação | 12 |
| 2.7.4 Migração Celular | 12 |
| 2.7.5 Organogênese | 12 |
| 2.8 A Diferenciação Celular | 13 |
| 3 Medidas em Redes Complexas | 15 |
| 3.1 Distribuição de Conectividades | 16 |

| | | |
|----------|--|-----------|
| 3.2 | Coeficiente de Aglomeração | 17 |
| 3.3 | O Menor Caminho Médio | 19 |
| 4 | Um Modelo para a Rede de Regulação Gênica | 21 |
| 4.1 | A Construção da Rede | 22 |
| 4.2 | A Caracterização da Rede | 24 |
| 4.2.1 | Distribuição de Conectividades | 25 |
| 4.2.2 | O Menor Caminho Médio | 26 |
| 4.2.3 | Coeficiente de Aglomeração | 27 |
| 4.2.4 | Modularidade | 29 |
| 5 | Um Modelo para a Dinâmica da Regulação | 32 |
| 5.1 | A Escolha da Regra | 33 |
| 5.2 | Atrator, Período e Transiente | 34 |
| 5.3 | Os Três Regimes de Comportamento | 36 |
| 5.4 | A Dinâmica da Rede é Independente do Grupo dos Genes Estruturais | 39 |
| 5.5 | O Tamanho dos Períodos e Transientes | 40 |
| 5.6 | O Número de Atratores | 43 |
| 5.7 | A Assimetria na Figura 5.2 | 45 |
| 6 | Um Modelo para a Diferenciação Celular | 47 |
| 6.1 | O Modelo | 48 |
| 6.2 | Resultados | 50 |
| 7 | Conclusões | 55 |
| 8 | Perspectivas Futuras | 57 |
| | Referências Bibliográficas | 58 |

Resumo

FUMIÃ, Herman Fialho, M. Sc., Universidade Federal de Viçosa, Fevereiro, 2010.
Um modelo para regulação gênica e a diferenciação celular nos estágios iniciais da embriogênese. Orientador: Marcelo Lobato Martins. Coorientadores: Silvio da Costa Ferreira Junior e José Arnaldo Redinz

A compreensão total do funcionamento das redes de regulação gênica provavelmente explicará a complexidade e a fisiologia dos organismos. O fenômeno da diferenciação celular, uma das conquistas da evolução que permite a existência de vida multicelular, também terá várias de suas peculiaridades elucidadas com a compreensão das redes de regulação. Neste trabalho, propomos um modelo para a topologia e a dinâmica de redes de regulação e, representando células por essas redes, estudamos como novos tipos celulares surgem através de divisões e interações entre várias células, ou seja, o fenômeno da diferenciação celular. Nossos resultados sugerem que a dinâmica de uma única célula pode ser congelada, marginal ou caótica. Contudo, somente a dinâmica marginal agrega conjuntamente estabilidade a mutações e capacidade evolutiva, características fundamentais à vida. Ainda mais, a diferenciação celular só acontece de maneira biologicamente aceitável quando temos células interagentes no regime marginal. Assim, o nosso modelo suporta a hipótese de que a vida acontece na “borda do caos”.

Abstract

FUMIÃ, Herman Fialho, M. Sc., Universidade Federal de Viçosa, February, 2010.
A model for gene regulation and cell differentiation in early stages of embryogenesis. Adviser: Marcelo Lobato Martins. Co-Advisers: Silvio da Costa Ferreira Junior and José Arnaldo Redinz.

The total understanding of genetic regulatory networks will likely explain the complexity and the physiology of organisms. The phenomenon of cell differentiation, one of the major transitions in evolution that allows the existence of multicellular life, will also have several of its peculiarities elucidated with such understanding of regulatory networks. In this work, we propose a model for the topology and dynamics of regulatory networks. Representing cells by these networks, we investigate how new cell types emerge from divisions and interactions between multiple cells, i.e., the phenomenon of cell differentiation. Our results suggest that the dynamics of a single cell can be frozen, marginal or chaotic. However, only the marginal dynamics exhibits both the stability to mutations and the ability to evolve, fundamental features of life. Moreover, cell differentiation only occurs in a biologically acceptable level only when the interacting cells are operating in the marginal regime. Thus, our model supports the hypothesis that life happens in the “edge of chaos”.

Capítulo 1

Introdução

A vida começou há muito tempo neste planeta, algo em torno de 3.8 bilhões de anos atrás. Embora os primeiros seres fossem muito rudimentares, ao longo de tantos anos a evolução produziu seres de estupenda complexidade. Os mais notáveis desses seres são, decerto, aqueles formados por milhares de células que trabalham harmoniosamente. Esses seres, chamados de multicelulares, podem ser dotados de capacidades refinadas de locomoção, obtenção e assimilação de energia, instinto social, entre outras. Entre eles encontram-se até mesmo nós, os únicos seres, ao que parece, complexos o suficiente para estarem cientes de sua própria existência e de seu fim inelutável.

Uma das características mais importantes de seres multicelulares, e que permite a própria existência desses organismos, é o fato deles serem formados por células de diferentes tipos. Esses tipos celulares distintos desempenham funções diversas nos organismos multicelulares, e é o seu trabalho conjunto que gera a complexidade observada nesses seres. Apesar disso, todo organismo multicelular começa de uma única célula, e é ao longo do seu desenvolvimento embrionário que os novos tipos celulares vão surgindo. Esse é justamente um dos objetivos deste trabalho: elaborar um modelo que seja capaz de gerar novos tipos celulares através de um processo

semelhante ao início do desenvolvimento embrionário. O outro objetivo, não menos importante, e que de fato nos permitirá a realização do primeiro, é a elaboração de um modelo para a regulação gênica, que parece ser uma das chaves para montar o quebra-cabeça da vida.

Neste trabalho, inicialmente definiremos alguns conceitos essenciais, tais como: genes, tipos celulares, DNA, a regulação gênica, desenvolvimento, entre outros. Isso será feito no Capítulo 2. Devido ao fato de que uma célula em nosso modelo será representada por uma rede de regulação gênica, no Capítulo 3 definiremos algumas medidas que são geralmente feitas para se caracterizar uma rede. No Capítulo 4, apresentaremos a rede de regulação proposta por nós e faremos a sua caracterização. No Capítulo 5, iremos propor uma dinâmica para essa rede de regulação gênica e comparar os resultados obtidos com aqueles para células de organismos reais. Nesse capítulo, trataremos ainda do estudo do funcionamento de uma única célula. Finalmente, no Capítulo 6 estudaremos a dinâmica de várias células interagindo. Estaremos especialmente focados no estudo do aparecimento de novos tipos celulares, o que ocorre como resultado dessas interações intercelulares. No Capítulo 7, apresentaremos a Conclusão de nosso trabalho e, no Capítulo 8, as Perspectivas Futuras.

Capítulo 2

Uma Breve Revisão Biológica

NESTE capítulo, serão apresentados alguns conceitos que são essenciais para a compreensão dos modelos para a regulação gênica e para a diferenciação celular que foram desenvolvidos neste trabalho. Inicialmente, discorreremos acerca de assuntos tais como DNA, genes e expressão gênica. Uma vez que esses conceitos-chaves tenham sido apresentados, abordaremos a biologia do desenvolvimento e da diferenciação celular.

2.1 DNA: a Molécula da Vida

O número de espécies de seres vivos diferentes no mundo é estimado entre 10 e 100 milhões[1]. Uma característica observada em todas essas espécies é que os “filhos” são sempre da mesma espécie dos “pais”. Embora isso pareça óbvio, uma vez que, por exemplo, não se viu até hoje uma égua gerar uma galinha, essa observação levanta a seguinte questão: Como estão armazenadas em cada espécie as informações que lhe permite gerar um ser “à sua imagem e semelhança”?

Hoje, sabemos a resposta para essa pergunta, e por mais estranho que possa parecer, quaisquer organismos vivos, desde amebas até seres humanos, armazenam

essas informações da mesma forma¹. Essa resposta se encontra no interior das células, para ser mais exato no núcleo celular². Lá, encontramos uma macromolécula chamada de **DNA** (ácido desoxirribonucléico). Tal qual uma grande biblioteca, é nessa molécula que estão armazenadas as informações para a construção de um ser vivo e também para o delicado processo de manutenção da vida. Essas informações se encontram codificadas através de diferentes formas: de maneira direta, codificada na estrutura do próprio DNA; e de maneira indireta, codificada na dinâmica e topologia do que chamamos de rede de regulação gênica. Antes de abordarmos esses conceitos, expliquemos a estrutura do DNA.

2.2 A Estrutura do DNA

A molécula de DNA é um polímero, isso significa que ela é um composto químico formado por unidades que se repetem. Essas unidades do DNA são chamadas de nucleotídeos e são formadas por: um açúcar desoxirribose, um grupo fosfato e uma base nitrogenada. A base nitrogenada pode ser de quatro tipos: adenina, citosina, guanina e timina; denotadas pelas letras **A**, **C**, **G** e **T**, respectivamente. Dessa forma, no DNA existem quatro tipos diferentes de nucleotídeos, que diferem entre si exclusivamente pelo tipo de base que possuem. Os nucleotídeos se encontram ligados uns aos outros através de uma ligação do grupo fosfato com o açúcar de outro nucleotídeo formando uma longa fita linear.

O DNA é formado por duas dessas longas fitas que se encontram enroladas entre si de maneira a formar uma estrutura de dupla hélice, que pode ser vista na figura 2.2. As duas fitas se unem através de ligações entre as bases dos nucleotídeos. Uma restrição imposta devido a esse tipo de ligação base-base é que nem todas as bases se ligam umas com as outras. Se um nucleotídeo de uma fita tiver a base A, ele só poderá

¹Esse fato é um forte indício de que todos os organismos descendem de um único ancestral comum.

²Uma ressalva precisa ser feita no caso das hemácias nos mamíferos. Durante o processo através do qual são geradas elas perdem o seu núcleo com o objetivo de obter mais espaço para o transporte de oxigênio.

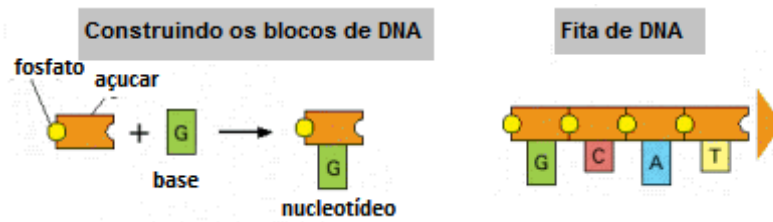


Figura 2.1: À esquerda, temos a estrutura dos nucleotídeos e, à direita, a maneira como eles se unem para formarem uma cadeia linear (observe que a ligação se dá através do fosfato com o açúcar). Figura adaptada de [1]

estar ligado a um nucleotídeo da outra fita que tenha a base T, e, similarmente, a base C só se liga com a G. Dessa forma, podemos dizer que as cadeias são complementares, ou seja, se soubermos a estrutura de uma delas necessariamente saberemos a da outra. Nessas seqüências de diferentes bases é que se encontram codificadas de forma direta várias das informações de um organismo.

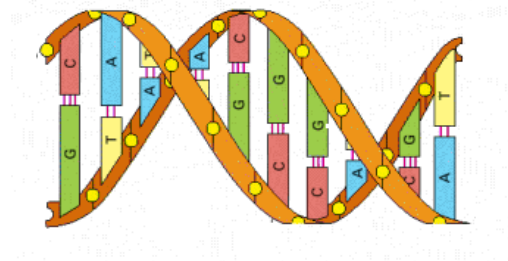


Figura 2.2: Estrutura em dupla hélice da molécula de DNA. Figura retirada de [1]

Entretanto, resta-nos explicar como a informação codificada nas bases do DNA é lida e interpretada pela célula.

2.3 Transcrição e Tradução: a Execução do Código

Se escolhêssemos aleatoriamente uma seqüência de nucleotídeos no DNA humano e tentássemos descobrir qual o papel daquela informação no funcionamento celular, a maior probabilidade é que ela não significasse nada. De fato, cerca de 95% do DNA

não guarda nenhuma informação que seja utilizada pela célula³. O DNA assemelha-se a um livro em que de dez páginas apenas uma estivesse escrita em alguma linguagem, e nas demais tivéssemos apenas letras dispostas aleatoriamente. Entretanto, a situação no DNA ainda é um pouco mais complicada. As informações que realmente significam alguma coisa não se encontram todas juntas, na verdade elas se encontram intercaladas entre seqüências de nucleotídeos que não significam nada. A situação é análoga a que temos na seguinte frase: ABFOE**ABOLA**JRNWECASAMCSS.

A cada um dos pedaços espalhados pelo DNA que codificam alguma informação utilizada pela célula damos o nome de **gene**. Podemos dizer de maneira simples que genes são *partes do DNA onde estão codificadas as informações necessárias para a produção de uma ou várias proteínas*.

As proteínas são o produto final do processo de execução da informação genética, e é através delas que praticamente todos os processos necessários a vida são levados a cabo. Elas são responsáveis, entre outras coisas, pela manutenção das estruturas celulares, geração de movimentos, sensibilidade a sinalização, regulação gênica. Em suma, quase tudo que ocorre nos organismos. Assim sendo, ao controlar a síntese de proteínas na célula os genes estão ditando grande parte do comportamento celular. Ainda nos resta responder como a informação dos genes se transforma em proteínas. Isso acontece mediante dois processos, que chamamos de **transcrição** e **tradução**. Na fase de transcrição, a informação contida em um dado gene é copiada para uma fita de **RNA**; e na tradução, a informação contida nesse RNA é utilizada para a produção da proteína que ela está codificando. O RNA é um polímero semelhante ao DNA, com a diferença de no seu nucleotídeo haver o açúcar ribose, ao invés da desoxirribose, e a base Timina (T) ser substituída pela base Uracila (U). Outra diferença do DNA é que o RNA só possui uma única fita. O processo de transcrição inicia-se quando a enzima **RNA polimerase** se liga a uma região do gene que vai ser transcrito chamada de **promotora**. A região promotora é necessária porque é ela quem indica onde começa o gene que vai ser transcrito. Entretanto, para que a RNA polimerase

³Esse tipo de DNA é chamado de DNA “lixo”, sendo esse termo usado porque não codificam nenhum tipo de proteína ou exercem qualquer função conhecida.

se ligue a região promotora é necessária a ajuda de proteínas chamadas de **fatores de transcrição**. Essas proteínas estão associadas ao controle da expressão gênica e elas têm o poder de ativar ou inibir o processo de transcrição. Uma vez que a RNA polimerase se ligue a região promotora, ela começara a construção de uma fita de RNA utilizando o gene como modelo. Quando o processo de transcrição termina, a fita de RNA resultante é levada para uma organela celular chamada de **ribossomo**. No ribossomo, a informação contida na fita de RNA será traduzida e utilizada para a produção de uma proteína. Se a proteína associada a um dado gene estiver sendo sintetizada na célula em um dado momento, dizemos que o gene está sendo **expresso**.

2.4 As Informações Contidas na Rede de Regulação Gênica

Utilizando tijolos, podemos fazer construções tão distintas como uma casa ou a Catedral de Notre-Dame, tudo depende do plano de construção. As recentes evidências sugerem que as informações contidas nos genes assemelham-se a tijolos, e ao que parece, os genes não guardam os planos de construção, pelo menos, não de maneira direta. De fato, o estudo experimental e teórico de genomas nos últimos anos evidencia que a complexidade dos organismos não é resultado do seu número de genes, mas sim do intrincado mecanismo de regulação entre eles. De acordo com esse novo paradigma, as características fenotípicas dos organismos não estão codificadas somente nas bases de cada gene, mas também na topologia da rede de regulação gênica e na dinâmica da mesma[2]. O plano de construção, ou seja, aquilo que nos explicará como determinar a estrutura, o funcionamento e a complexidade dos seres vivos, só poderá ser encontrado no estudo da topologia e da dinâmica de toda a rede de regulação gênica.

2.5 A Expressão Gênica e os Diferentes Tipos Celulares

Observe os dois grupos de células da figura (2.5). À esquerda, nós temos neurônios, que são células do sistema nervoso; e à direita, temos células musculares cardíacas⁴, que entram na formação do miocárdio⁵. A diferença estrutural entre os dois tipos de células é gritante e a diferença funcional também. Enquanto os neurônios estão associados à transmissão de impulsos elétricos, as células musculares do coração são responsáveis pelas contrações que permitem esse órgão bombear o sangue através do organismo. Ninguém discordará que esses dois tipos de células são fundamentais, mas

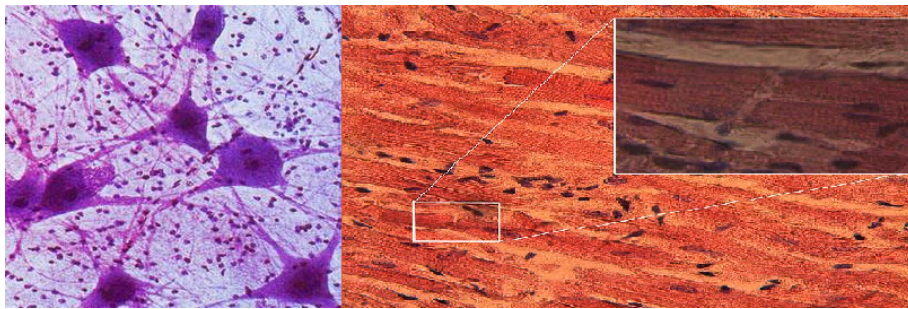


Figura 2.3: À esquerda temos neurônios e à direita células musculares do tecido cardíaco. Figuras retiradas de: neurônios [3] e células cardíacas[4].

o que realmente difere essas células a nível genético? A resposta é o seu padrão de **expressão gênica**. Por padrão de expressão gênica estamos nos referindo ao conjunto de todos os genes que estão sendo expressos em um dado momento. Padrões de expressão distintos implicam a produção de diferentes tipos de proteínas, e é devido a isso que as células são diferentes, tanto funcional quanto estruturalmente. A existência de diferentes tipos celulares agindo de maneira cooperativa é um pré-requisito para a existência dos organismos multicelulares. Sem ela o mundo seria um lugar “sem graça”, formado tão somente por bactérias, amebas e outros organismos unicelulares.

⁴Existem três tipos de células musculares, que entram na formação dos três tipos de tecidos musculares existentes, a saber: tecido muscular liso (paredes do intestino, vasos sanguíneos, etc.), tecido muscular estriado (músculos) e tecido muscular cardíaco (coração).

⁵O miocárdio é a parte média da parede do coração e é responsável pelos batimentos cardíacos.

2.6 A Biologia do Desenvolvimento

O mistério da reprodução tem intrigado a humanidade desde as épocas mais primevas. Embora todas as culturas tenham dado as suas explicações, o primeiro tratamento científico do problema que se têm notícia foi feito por Hipócrates⁶, na Grécia, no século V a.C. Usando as idéias correntes daquele tempo, ele tratou de explicar o desenvolvimento em termos de princípios de calor, umidade e solidificação. Entretanto, foi Aristóteles⁷, importantíssimo filósofo grego, quem levantou a questão que perduraria até o final do século XIX: Como as diferentes partes do embrião são formadas? Aristóteles propôs duas possibilidades: uma era de que tudo no embrião já estaria pré-formado desde o início e apenas se tornava maior durante a gestação, esse ponto de vista ficou conhecido como **pré-formacionismo**. A outra possibilidade era de que as novas estruturas iam surgindo gradualmente ao longo do percurso do desenvolvimento embrionário, em um processo que ele denominou de **epigênese** (que quer dizer “no curso da formação”). Aristóteles inclinava-se pela epigênese, e, como sabemos hoje, ele estava correto. A influência de Aristóteles no pensamento europeu foi enorme, estendendo-se desde a “cristianização” de suas idéias por São Tomás de Aquino no século XIII até boa parte do século XVII. Devido a isso, durante esse período a epigênese era a idéia mais defendida. Entretanto, a partir do final do século XVII o pré-formacionismo voltou a ser propalado. Muitos não acreditavam que forças físicas ou químicas pudessem moldar um ser vivo. Além disso, a aparição do microscópio revelou a existência dos espermatozóides, que foram logo interpretados como os “animais” cujo crescimento daria lugar ao organismo completo. Havia mesmo quem afirmasse que era possível se enxergar um “homenzinho” na cabeça do espermatozóide. Por último, o desenvolvimento e aplicação do cálculo integral permitiram a suposição de que a matéria era divisível até o infinito e, portanto, poderiam existir estruturas extremamente pequenas no espermatozóide que cresceriam durante o de-

⁶Hipócrates, conhecido como o pai da medicina, foi uma das figuras mais importantes na história da saúde. É, sobretudo, conhecido pelo famoso Juramento de Hipócrates.

⁷Pela amplitude dos campos em que trabalhou e pelo rigor de sua metodologia, Aristóteles é considerado por muitos como o primeiro pesquisador científico no sentido atual do termo.

envolvimento embrionário. A discussão epigênese/pré-formacionismo continuou até o desenvolvimento da teoria celular entre 1820 e 1880 pelo botânico alemão Mathias Schleiden e pelo fisiologista Theodor Schwann. As células foram reconhecidas como as unidades básicas formadoras dos organismos vivos e foi descoberto que células surgem somente da divisão de outras células. Os organismos, como plantas e animais, passaram a ser vistos como comunidades celulares. Em face dessas novas descobertas, a epigênese tornou-se uma opção muito mais provável, ficando assim encerrado o dilema epigênese/pré-formacionismo.

2.7 Etapas do Desenvolvimento Embrionário

Quando o gameta masculino fecunda o gameta feminino dando origem a uma célula chamada de **zigoto**, inicia-se uma série de processos complexos de divisão, migração, sinalização e diferenciação celular, que culminam na formação de um novo organismo⁸. Entender esses processos tem sido uma empreita corajosa da biologia e podemos dizer que temos uma compreensão razoável dele no estágio atual. Embora a muito não olhemos o desenvolvimento embrionário como miraculoso, torna-se difícil não ficar estupefato com ele. A título de ilustração, neste tópico apresentaremos de maneira resumida as etapas básicas do desenvolvimento embrionário em vertebrados. Vale ressaltar que essas etapas apresentadas não são as mesmas para os outros organismos (plantas, insetos, etc.), apesar de guardarem semelhanças em vários aspectos.

2.7.1 Clivagem

Como já dissemos, a fecundação do óvulo (gameta feminino) pelo espermatozóide (gameta masculino) dá origem a uma única célula fertilizada que chamamos de **zigoto** ou **célula ovo**. Dentro de algumas horas após a fertilização, o zigoto começa a se

⁸Nos animais os gametas masculinos e femininos chamam-se respectivamente, espermatozóide e óvulo. Nas plantas chamam-se respectivamente, anterozóide e oosfera.

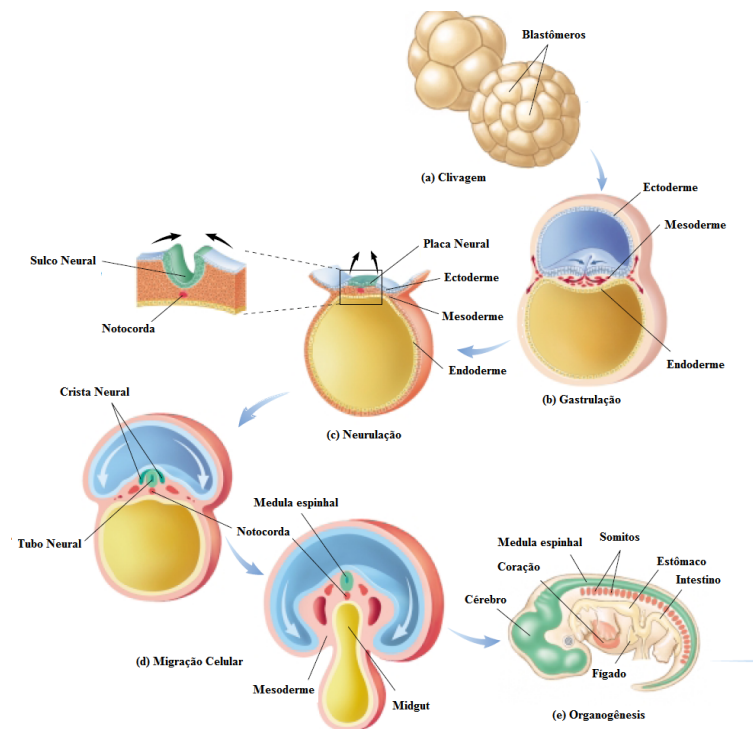


Figura 2.4: Uma ilustração dos principais eventos do desenvolvimento. Figura adaptada de [5].

dividir rapidamente em células cada vez menores, até que tenhamos uma esfera formada por várias células diminutas. A esfera final é praticamente do mesmo tamanho do zigoto, apesar de ser formada por centenas de células. A essas células diminutas denominamos **blastômeros**.

2.7.2 Gastrulação

Quando a clivagem chega ao fim, a taxa com a qual as células se dividem se torna bem menor. Os blastômeros iniciam uma série de movimentos onde mudam as suas posições uns em relação aos outros. Essa série de rearranjos dos blastômeros é chamada de gastrulação. Como resultado final dessa fase nós temos a formação de três tipos de folhetos germinativos: a ectoderme, a endoderme e a mesoderme. Podemos entender esses folhetos germinativos como conjuntos de células que possuem destinos finais na formação dos tecidos e órgãos já determinados, que são os seguintes:

- A ectoderme dá origem a epiderme e ao sistema nervoso.
- A endoderme forma o revestimento do intestino e do sistema respiratório.
- A mesoderme dá origem aos rins, coração, gônadas, e células sanguíneas.

2.7.3 Neurulação

Logo depois que a gastrulação está completa, uma grande parte da ectoderme na superfície dorsal do embrião começa a tornar-se espessa. O espessamento é produzido pelo alongamento de certas células ectodérmicas. Essas células assumem o formato de uma cunha por meio da contração de feixes de actina e posteriormente enrolam-se formando um tubo. Este tubo é chamado de **tubo neural** e é a partir dele que a medula espinhal e o cérebro serão formados [5].

2.7.4 Migração Celular

Durante essa fase as células começam a se movimentar no embrião com o objetivo de alcançarem sua posição correta, por assim dizer. Células de mesmo tipo tendem a ficarem próximas para formarem um tecido. É através de proteínas receptoras na sua superfície que uma célula "sabe" se deve permanecer em um dado lugar ou continuar migrando.

2.7.5 Organogênese

Logo que os folhetos germinativos estão formados, as células começam a interagir umas com as outras de modo a formar os órgãos e os tecidos. Essa fase é chamada de organogênese. Vários processos atuam concomitantemente nessa etapa de maneira a permitir essa formação de novas estruturas, como por exemplo, sinalização celular, migração celular e adesão celular diferenciada.

2.8 A Diferenciação Celular

Conforme já comentamos, organismos multicelulares somente são possíveis devido à diferenciação celular. Graças a ela, existem células diferenciadas capazes de executar as mais distintas funções, que são fundamentais para o funcionamento de organismos complexos como nós, por exemplo. Já foi dito também que, durante o desenvolvimento embrionário, iniciado a partir do zigoto, nós observamos passo a passo a formação dos órgãos. Obviamente, isso implica que as células estão se diferenciando, uma vez que os órgãos são formados de células que têm características e funções distintas. Pretendemos explicar agora, quais tipos de processos permitem o aparecimento de células diferentes durante o desenvolvimento.

São dois os mecanismos por intermédio dos quais as células conseguem se tornar diferentes umas das outras. O primeiro deles, que ocorre em alguns organismos como o nematódeo *Caenorhabditis elegans* e a mosca da fruta *Drosophila melanogaster* entre outros, é chamado de **divisão assimétrica**. Nesses organismos, a entrada do espermatozóide no óvulo condiciona uma reorientação das moléculas, de modo que algumas zonas podem estar mais ricas ou com escassez de um dado tipo de molécula. Entre essas moléculas temos algumas que estão associadas ao controle da regulação gênica, como os fatores de transcrição. Quando a célula ovo se divide, as duas novas células formadas adquirem quantidades diferentes dos fatores de transcrição e outras moléculas associadas à regulação. Conseqüentemente, as duas células possuem padrões diferentes de expressão gênica, ou seja, já se tornaram diferenciadas.

Entretanto, a maior parte dos organismos não apresenta divisão assimétrica e mesmo naqueles que apresentam, o segundo mecanismo de diferenciação exerce um papel importante. Chamamos esse processo de **diferenciação por indução**. Nele, o controle do destino de uma célula é exercido pelas outras células do organismo através de sinalização. Essa sinalização pode ser feita a partir de moléculas que são difusíveis no meio, tendo portanto um alcance de atuação mais amplo, ou por contato direto entre as células. Seja como for, o efeito final é que essas moléculas acabam por atuar diretamente no interior da célula alterando o seu padrão de expressão gênica e, logo,

influenciando o seu destino celular.

Capítulo 3

Medidas em Redes Complexas

UMA rede pode ser definida como um conjunto de elementos que se encontram relacionados (ligados) de alguma forma. Com uma definição tão ampla, podemos encontrar estruturas do tipo rede nos lugares mais inusitados. Vamos supor, por exemplo, uma festa. Podemos criar uma rede estabelecendo, por exemplo, que cada pessoa será um elemento da rede, e que haverá uma ligação entre um elemento e outro se as pessoas que eles representam se conhecerem. Outros exemplos possíveis seriam: uma rede onde os nós fossem as cidades de uma dada região e as ligações entre os nós a existência de uma estrada ligando uma cidade à outra; a internet, onde cada página da rede está conectada a várias outras por meios de *links*, ou ainda as redes de atores, onde cada ator está conectado a outro se eles já tiverem atuado em um mesmo filme. Todas essas redes são exemplos do que chamamos de redes complexas, assim chamadas por que as maneiras como os nós se conectam não exibem nenhum tipo de padrão simples. Devido a essa dificuldade na determinação da maneira como os nós se conectam, a primeira hipótese elaborada foi a de que essas conexões eram estabelecidas ao acaso. Essa hipótese, proposta por Paul Erdos e Alfred Rényi, permaneceu por bastante tempo. Entretanto, com o crescimento do número de informações sobre redes complexas acumuladas durante os últimos anos, e também devido ao aumento da capacidade computacional, mostrou-se que as redes reais não são completamente

aleatórias [6]. De fato, várias medidas realizadas nessas redes produziram resultados distintos daqueles que seriam esperados para uma rede aleatória. Essas medidas se tornaram ferramentas indispensáveis para a caracterização de redes complexas. Neste capítulo, o nosso objetivo é unicamente definir algumas dessas grandezas fundamentais. Isso se faz necessário porque no próximo capítulo precisaremos caracterizar a rede que iremos propor.

3.1 Distribuição de Conectividades

Chamamos os elementos da rede que se conectam uns com os outros de **vértices** ou **nós**, e as ligações de **arestas** ou **arcos**. Cada vértice i em uma rede possui um número k_i de arestas. Chamamos esse número de conectividade ou grau do vértice. Dizemos que uma rede é **direcionada** se as arestas que ligam os vértices possuem uma direção. Por isso, para redes direcionadas cada vértice i tem dois tipos de conectividade: uma que mede o número de arestas direcionadas do vértice i para os outros vértices (k_{out} , ou seja, saindo do vértice), e uma que mede o número de arestas vindas de outros vértices para o vértice i (k_{in} , ou seja, entrando no vértice). Em redes **não-direcionadas** não existem direções entre as arestas. Conseqüentemente, só temos um tipo de conectividade.

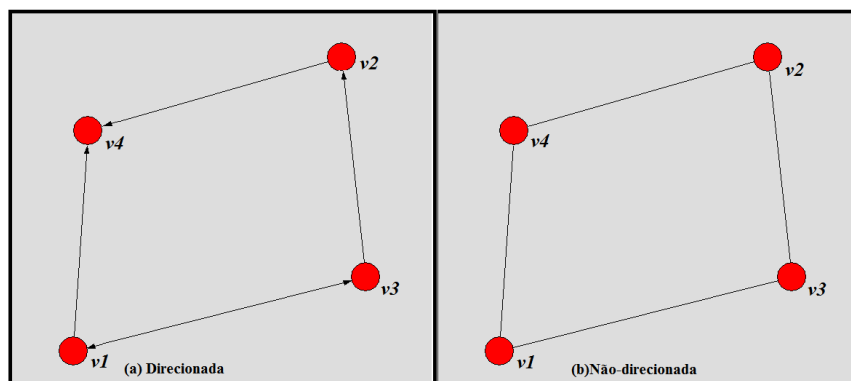


Figura 3.1: (a) Rede direcionada formada por 4 vértices, onde os pontos vermelhos correspondem aos vértices, as linhas às arestas e as setas indicam a direção da ligação. (b) Rede não-direcionada correspondente àquela da figura 3.1(a).

Observando a figura 3.1, podemos entender esses conceitos mais claramente. A figura 3.1(a) é uma rede direcionada, e a figura 3.1(b) a sua correspondente não-direcionada. Para a rede direcionada, o vértice v_3 , por exemplo, possui $k_{out} = 2$, uma vez que existem duas arestas saindo dele: uma que aponta para o vértice v_1 e outra para o vértice v_2 . Diferentemente, o vértice v_3 possui $k_{in} = 1$, uma vez que temos apenas uma aresta, apontando do vértice v_1 para ele. Para a rede não-direcionada, o vértice v_3 possui uma conectividade $k = 2$, já que existem duas arestas que o ligam a v_1 e v_2 .

Redes direcionadas e não-direcionadas carregam informações diferentes. Nas redes direcionadas, o fato de um vértice i afetar um vértice j de alguma forma, não implica necessariamente que o vértice j também afete o vértice i . Nas redes não-direcionadas, entretanto, isso é sempre verdade, e, portanto, não existe necessidade de se dar direção às arestas. Portanto, o uso de redes direcionadas ou não-direcionadas dependerá do problema ao qual queiramos abordar. Redes de regulação gênica, por exemplo, são melhor modeladas por redes direcionadas, já que o fato de um gene i controlar a expressão de um gene j não implica que a recíproca seja verdadeira.

Na maior parte das vezes, quando estamos estudando uma rede não estamos interessados na conectividade dos vértices individualmente, e sim no que chamamos de distribuição de conectividades $P(k)$. Essa medida nos informa qual a probabilidade de um vértice i possuir um grau k_i . Redes aleatórias, por exemplo, possuem uma distribuição de conectividades do tipo Poisson com um pico na conectividade média dos vértices. Essa é uma das medidas que permitiu a verificação de que redes complexas reais não eram aleatórias. Para várias redes reais a distribuição de conectividades encontrada era do tipo lei de potência ou exponencial, e não poissoniana.

3.2 Coeficiente de Aglomeração

Uma característica notada na maioria das redes reais é que os vértices tendem a criar grupos que são muito conectados entre si. Um exemplo disso seria uma rede

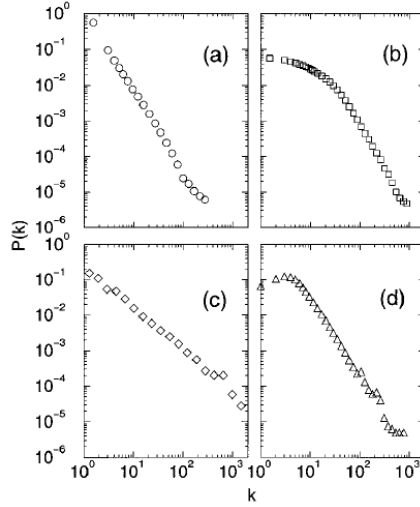


Figura 3.2: Distribuição de conectividade de algumas redes reais: (a)Internet no nível de roteadores; (b)rede de co-atuação de atores; (c)rede de co-autoria de físicos de alta-energia;(d)rede de co-autoria de neurocientistas. Figura retirada de [6]

formada por laços de amizade, onde existe uma grande chance de vários amigos serem amigos entre si. O coeficiente de aglomeração é uma quantificação de quão coeso um determinado grupo de vértices está. Ele é definido para um vértice i , como a razão entre o número de arestas entre os vizinhos desse vértice, denotado por A_i , e o número máximo possível de arestas entre esses vizinhos, dado por $(k_i(k_i - 1))/2$. Em geral, estamos interessados no coeficiente de aglomeração médio da rede, dado pela expressão a seguir:

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^N c_i = \frac{1}{N} \sum_{i=1}^N \frac{2A_i}{k_i(k_i - 1)} \quad (3.1)$$

em que N é o número de vértices da rede. Como vemos, o coeficiente de aglomeração médio nada mais é do que a média do coeficiente de aglomeração para cada vértice (c_i).

Na figura 3.2 são mostrados exemplos de grafos ou redes com valores distintos de coeficiente de aglomeração associado ao vértice central que possui a mesma conectividade.

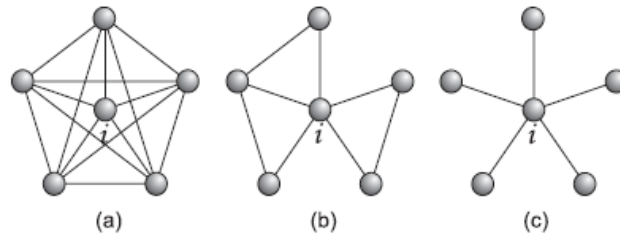


Figura 3.3: Exemplos do valor do coeficiente de clustering para três redes. (a) Os vértices estão totalmente conectados, $\langle C \rangle = 1$. Em (b), $\langle C \rangle = 3/10$. Em (c), $\langle C \rangle = 0$, já que os vizinhos não possuem conexões entre si. Figura retirada de [7]

3.3 O Menor Caminho Médio

A distância que separa um vértice de outro é uma informação importante em redes complexas. Para a rede da Internet, por exemplo, a perda de pacotes trocados entre computadores pode aumentar com a distância entre eles, já que quanto maior a distância, maior o número de roteadores no caminho e, conseqüentemente, maior a probabilidade de ocorrer falhas ou congestionamentos [7].

O menor caminho que leva de um vértice a outro é, de todos os caminhos possíveis, aquele pelo qual precisamos passar pelo menor número de arestas para chegar ao vértice desejado. Na figura 3.3, por exemplo, o menor caminho que liga o vértice v_1 ao vértice v_5 é aquele que faz o trajeto $v_1-v_3-v_5$, e, portanto, o menor caminho é 2.

Geralmente, estaremos interessados no menor caminho médio. Para calculá-lo, temos que tomar a média sobre o menor caminho entre todos os pares de vértices i e j da rede.

Um fenômeno interessante, presente em muitas redes reais, é o chamado efeito mundo pequeno (*“small world”*). Redes com esse efeito podem apresentar milhares de vértices e, apesar disso, ter um menor caminho médio muito pequeno. A mais popular manifestação do efeito mundo pequeno é o conceito dos “seis graus de separação”, descoberto pelo psicólogo social Stanley Milgram (1967), que concluiu que em média

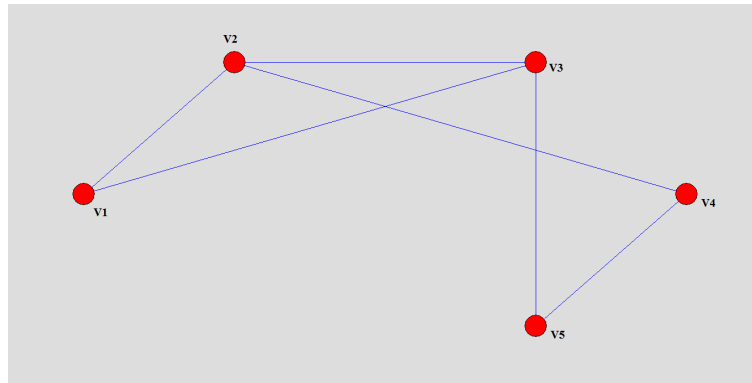


Figura 3.4: Rede não-direcionada formada por 5 vértices. Aqui utilizada com o intuito de ilustrar o conceito de menor caminho.

quaisquer duas pessoas nos EUA estão separadas através de laços de familiaridade por seis outras pessoas [6].

Capítulo 4

Um Modelo para a Rede de Regulação Gênica

NESTE capítulo, apresentaremos a primeira parte de nosso trabalho, que foi a construção de uma rede de regulação gênica que se aproximasse o máximo possível das redes reais de regulação. Essa fase do trabalho é essencial devido ao fato de que no estado atual do conhecimento biológico não temos nenhuma rede de regulação que esteja completamente determinada. Contudo, muitas características compartilhadas por essas redes são conhecidas e, se não podemos fazer o estudo de uma rede específica, o que representaria uma determinada espécie, podemos construir um *ensemble* de redes que englobem essas características gerais. Uma vez que isso seja feito, poderemos estudar a dinâmica regulatória intracelular e a diferenciação. Esperamos com isso observar fenômenos que apareçam como resultado dessas características comuns e se perdemos em especificidade por um lado, talvez ganhemos em conhecimento dos mecanismos básicos gerais por outro.

4.1 A Construção da Rede

Para a construção da rede de regulação gênica nós fomos guiados por quatro características que são observadas nesse tipo de rede, a saber:

1. **A existência de dois grupos de genes.** Os genes podem ser classificados em dois grupos: o grupo dos genes reguladores, que é formado por genes que controlam a expressão de outros genes (inclusive de outros genes reguladores), e o grupo dos genes estruturais, que é formado por genes que codificam proteínas que entram na formação de estruturas celulares ou têm função enzimática [8].
2. **A direcionalidade das redes.** As redes de regulação são redes direcionadas. Isso significa que se um gene i regula a expressão de um gene j , não necessariamente o gene j regulará a expressão do gene i .
3. **A distribuição de conectividades.** Como as redes de regulação gênica são direcionadas, nós temos a distribuição das conectividades que estão saindo (k_{out}) e das que estão chegando aos nós (k_{in}). Podemos dividir essas conexões ainda de duas formas: aquelas que são feitas somente entre os genes reguladores, e aquelas que os genes reguladores fazem com os estruturais. Considerando somente as ligações entre os genes reguladores, sabemos que a distribuição de conectividades k_{out} é do tipo lei de potência e a distribuição de conectividades k_{in} é do tipo Poisson [9]. No caso das conexões que são feitas entre os genes estruturais e os reguladores, a distribuição de conectividades é do tipo Poisson, sendo do tipo k_{in} para os genes estruturais, uma vez que eles só recebem ligações; e do tipo k_{out} para os reguladores, uma vez que eles só enviam ligações. Note que não existe uma distribuição de conectividades entre os genes estruturais, já que esses genes somente podem ser regulados e nunca reguladores.
4. **A regulação da expressão gênica.** Um gene pode regular o outro e também a si mesmo (auto-regulação) de duas maneiras: ativando ou inibindo a expressão gênica.

Tendo em vista essas características, nós construímos a rede da seguinte maneira:

- Começando com uma rede com N genes (nós), nós a dividimos em dois grupos. O tamanho dos grupos é determinado pelo parâmetro chamado *perc*, que pode variar de 0 a 1. Se *perc* vale 0.3, por exemplo, isso significa que 30% dos N genes pertencem ao grupo dos genes reguladores e os outros 70% ao grupo dos genes estruturais (**Característica 1**).
- No grupo dos genes reguladores, cada gene i regulará k_i outros genes, escolhidos aleatoriamente, que também sejam pertencentes ao grupo dos genes reguladores. Esse número k_i é escolhido de acordo com uma distribuição de probabilidades em lei de potência com expoente γ . Dessa forma, a distribuição de conectividades k_{out} entre os genes reguladores será do tipo lei de potência com expoente γ . O fato de escolhermos aleatoriamente com mesma probabilidade os genes que cada gene i regulará, faz com que a distribuição de conectividades k_{in} seja do tipo Poisson (**Características 2 e 3**).
- Para cada um dos genes j do grupo dos genes estruturais, um número k_j é escolhido de acordo com uma distribuição do tipo Poisson com conectividade média $\langle k \rangle$. Cada um dos genes j escolhe aleatoriamente com a mesma probabilidade k_j genes do grupo dos genes reguladores e recebe ligações deles, ou seja, são regulados por estes genes (**Características 2 e 3**).
- Em nosso modelo, consideramos que todos os genes reguladores se auto-regulam¹. Por último, resta-nos atribuir pesos entre as ligações, que vão indicar se uma dada conexão é inibidora da transcrição ou ativadora. Admitimos que toda ligação reguladora tem a mesma intensidade, podendo valer 1 ou -1 , conforme ela seja ativadora ou inibidora, respectivamente. O peso das ligações é escolhido aleatoriamente com probabilidade $P1$ de a ligação ser inibidora e $(1 - P1)$ de ser ativadora. (**Característica 4**).

¹Isso é obviamente uma simplificação em nosso modelo. Em parte, isso foi assumido para que haja uma diminuição do espaço dos parâmetros e, também, devido ao percentual de genes que se auto-regulam não ser uma informação fácil de se encontrar na literatura.

Um exemplo típico de rede de genes gerada conforme o procedimento discutido anteriormente é mostrado na figura 4.1.



Figura 4.1: Exemplo de rede de regulação gerada pelo nosso modelo. Os genes são representados pelas bolinhas e as setas entre dois genes indicam que um gene é regulado (seta apontando para ele) ou regula o outro gene (seta apontada para outro gene). Os dois grupos de genes estão misturados na rede. Um gene que somente possua ligações com setas apontando para ele é um gene estrutural. A auto-regulação não é apresentada na figura.

4.2 A Caracterização da Rede

Passemos agora à caracterização das propriedades exibidas pela rede proposta e sua comparação com os resultados apresentados na literatura para as redes de regulação gênica reais.

4.2.1 Distribuição de Conectividades

Embora já tenhamos proposto nossa rede de forma a obedecer a uma distribuição de conectividades similar à observada em redes reais, é necessário demonstrar que efetivamente isso foi feito de maneira correta.

Na figura 4.2, temos a distribuição de conectividades k_{out} do grupo dos genes reguladores para diferentes expoentes γ . Observamos claramente uma distribuição do tipo lei de potência, como era o esperado.

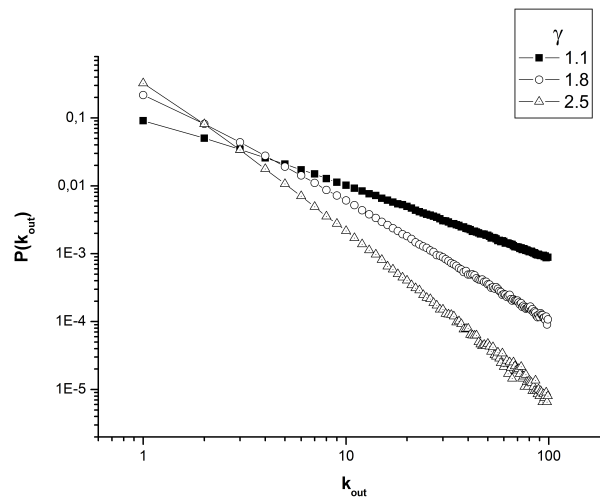


Figura 4.2: Distribuição de conectividades para k_{out} do grupo dos genes regulatórios. Média sobre 10.000 redes com $N = 100$ genes reguladores.

Na figura 4.3, temos a distribuição k_{in} correspondente a figura 4.2. Observamos uma distribuição do tipo Poisson centrada no valor da conectividade média da rede. A conectividade média da rede é dependente do valor de γ e do número de sítios que a rede possui. Na figura 4.4, temos a distribuição de conectividades k_{in} do grupo dos genes estruturais. Como dissemos, essa distribuição é escolhida de acordo com uma distribuição de Poisson com uma determinada conectividade média. Na figura 4.4, podemos ver que a escolha de $\langle k \rangle$ é totalmente independente do valor que tomamos para o expoente γ , como deveria efetivamente ser, posto que as ligações entre os genes reguladores e estruturais são escolhidos de maneira independente daquelas entre os

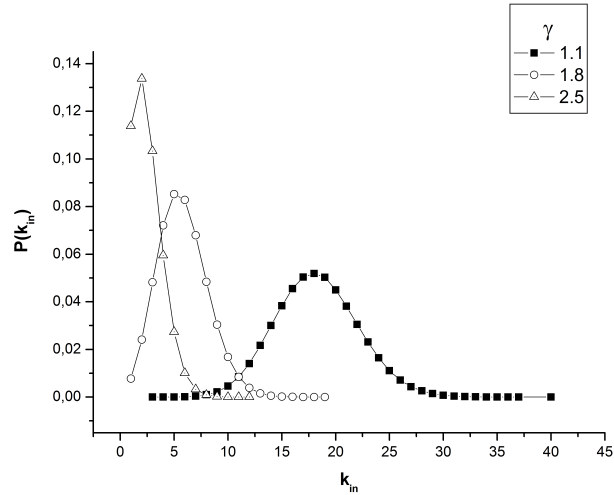


Figura 4.3: Distribuição de conectividades para k_{in} do grupo dos genes regulatórios. Média sobre 10.000 redes com $N = 100$ genes.

genes reguladores.

4.2.2 O Menor Caminho Médio

As redes de regulação, assim como a maior parte das redes reais, exibem o efeito mundo pequeno, ou seja, possuem um menor caminho médio que é pequeno se comparado ao número de nós da rede. Essa é uma propriedade importante em redes de regulação porque está associada a uma rápida transferência de informação através da rede.

Para o cálculo do menor caminho médio teremos que desconsiderar o fato das redes serem direcionadas e iremos tratá-las como se fossem não-direcionadas. O motivo disso é que para redes direcionadas temos vários nós a partir dos quais não podemos atingir outros nós da rede, e, nesse caso, o menor caminho seria infinito. Poderíamos contornar isso utilizando uma definição alternativa, chamada de eficiência global, que nada mais é do que considerarmos o inverso do menor caminho [7]. Contudo, como o procedimento de ignorar a direcionalidade da rede é aquele utilizado para o cálculo do menor caminho médio nas redes reais, será esse que iremos utilizar.

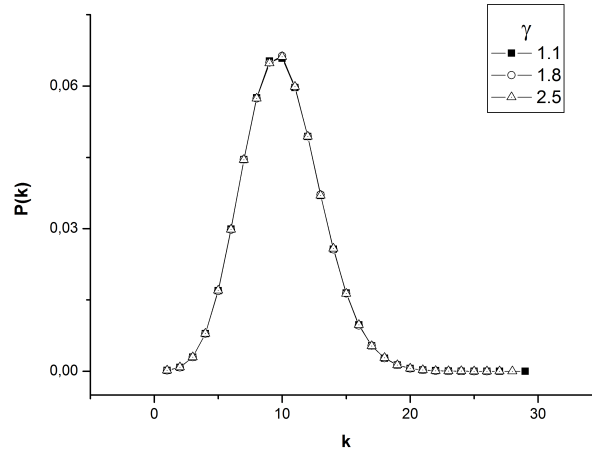


Figura 4.4: Distribuição de conectividades para k_{in} do grupo dos genes estruturais com $\langle k \rangle = 10$. Média sobre 10.000 redes com $N = 100$ genes estruturais e reguladores, $perc = 0.3$.

Como vemos da figura 4.5, o efeito mundo pequeno se evidencia em nosso modelo. Embora seja dependente do expoente γ , observamos que o menor caminho médio se mantém pequeno se comparado ao número de nós da rede, e aumenta muito lentamente com o aumento da mesma.

4.2.3 Coeficiente de Aglomeração

O coeficiente de aglomeração das redes regulatórias de organismos vivos pode variar consideravelmente dependendo da espécie a qual a rede pertence. Para o *Saccharomyces cerevisiae*, por exemplo, o coeficiente de aglomeração vale 0.11 [10], enquanto algumas redes podem ter um coeficiente tão elevado quanto 0.88 [11]. Apesar dessa grande variação, o fato é que para as redes de regulação o coeficiente de aglomeração é bem maior do que o esperado se a dita rede fosse uma rede aleatória. Outra propriedade observada é que o valor do coeficiente de aglomeração parece ser aproximadamente independente do tamanho da rede [9].

Para calcularmos o coeficiente de aglomeração para a nossa rede, utilizamos o

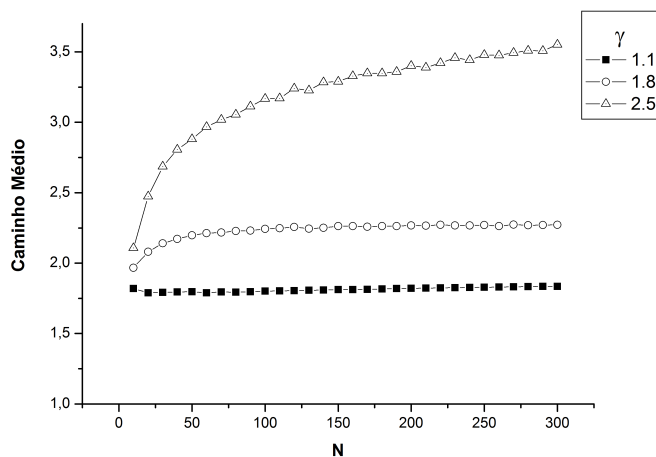


Figura 4.5: Menor caminho médio em função do tamanho da rede. Média sobre 1000 redes.

procedimento de ignorar a direcionalidade da mesma. Isso é feito porque a maior parte dos dados sobre redes complexas, de onde se conclui que elas geralmente possuem um alto coeficiente de aglomeração, é sobre redes não-direcionadas.

Como podemos ver da tabela abaixo, o coeficiente de clustering de nossa rede é consideravelmente maior do que aquele de sua versão aleatória equivalente.

| Expoente(γ) | C | $C_{aleat.}$ |
|----------------------|-------|--------------|
| 1.1 | 0.585 | 0.286 |
| 1.6 | 0.430 | 0.107 |
| 2.5 | 0.085 | 0.024 |

Na figura 4.6, verificamos que o coeficiente de clustering torna-se independente do tamanho da rede depois de uma pequena fase transiente. Observamos também que ele satura em um valor bem alto comparado com a versão aleatória correspondente, condizente com o que é observado em redes reais.

É importante ressaltar que em todas as considerações feitas acima, estávamos analisando somente o grupo dos genes reguladores (o que corresponde no nosso modelo a fazer $perc = 1$). Isso se deve ao fato de que as informações para redes reais

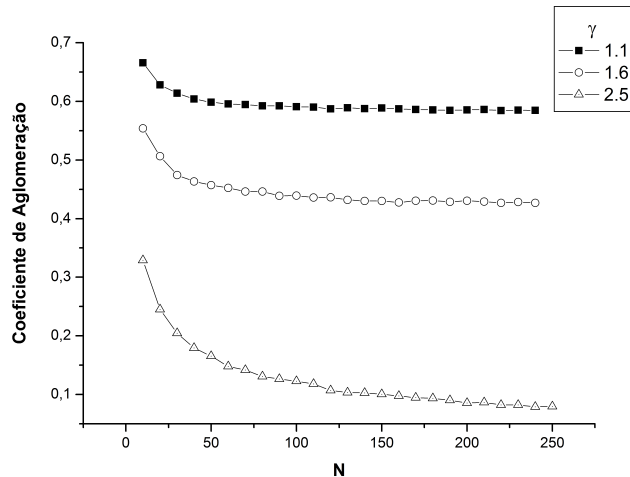


Figura 4.6: Variação do coeficiente de aglomeração com o tamanho da rede. Média sobre 2000 redes com até 250 nós.

apresentadas são baseadas somente na análise desse grupo.

4.2.4 Modularidade

Tem sido registrado na literatura que uma extensa variedade de redes exibe uma estrutura modular. Isso significa que se olharmos a rede como um todo poderemos identificar grupos de nós que estão muito conectados entre si e muito pouco com o restante da rede. Dessa forma, a rede seria formada por vários grupos ou módulos diferentes, formando uma estrutura semelhante a da figura 4.7(a).

Em redes modulares geralmente os integrantes de um mesmo módulo estão associados com um mesmo tipo de função na rede. Nós poderíamos esperar, por exemplo, que nas redes regulatórias, diferentes processos celulares fossem regulados por diferentes módulos. Realmente redes regulatórias apresentam módulos. Entretanto, o panorama é bem distinto daquele da figura 4.7(a). Na verdade, as redes regulatórias tendem a ser muito interconectadas e somente poucos módulos podem ser inteiramente separados do restante da rede [13]. A estrutura real dessas redes é semelhante a da 4.7(b).

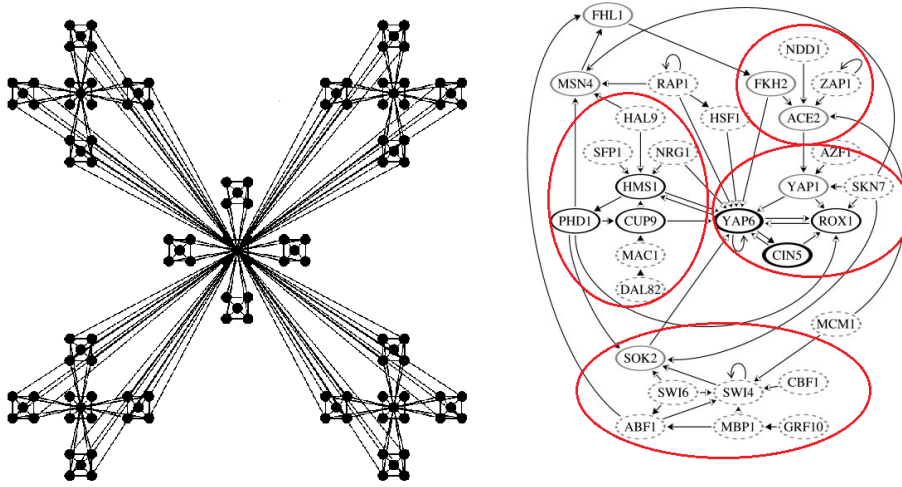


Figura 4.7: (a) Rede modular gerada a partir do modelo de Ravasz *et al.*[12].(b) Uma parte da rede de regulação da levedura *Saccharomyces cerevisiae*[13]

Ravasz *et al.* [12] mostraram que para várias redes reais livres de escalas o coeficiente de aglomeração (C) depende da conectividade do nó k de acordo com uma lei de potência $C \propto k^\alpha$. Esse comportamento foi associado por eles à presença de modularidade na rede. Com vistas a verificar se esse tipo de comportamento aparece em nossa rede, estudamos a variação do coeficiente de clustering com a conectividade (figura 4.8), e realmente verificamos uma dependência na forma de uma lei de potência, o que sugere a presença de estrutura modular em nossa rede.

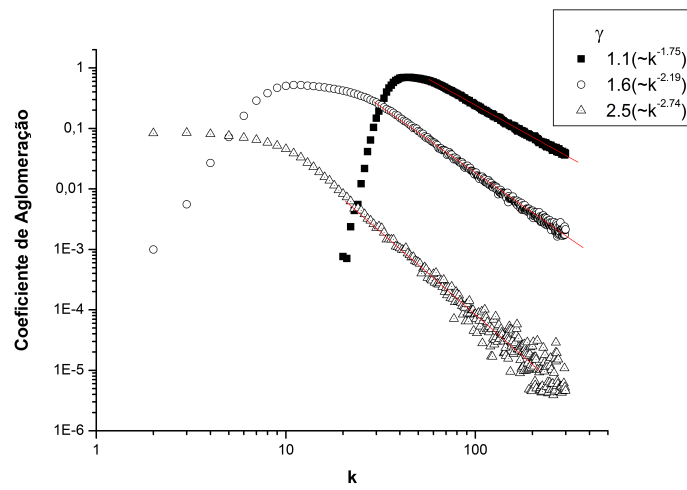


Figura 4.8: Coeficiente de aglomeração em função da conectividade do nó. Média sobre 500 redes para cada ponto, $perc = 1.0$, $N = 300$ genes.

Capítulo 5

Um Modelo para a Dinâmica da Regulação

COM os avanços recentes da biologia, temos obtido um painel cada vez mais completo sobre o processo de regulação gênica. Uma das técnicas mais importantes nessa área é a baseada no uso de *microarrays* de DNA. Essa técnica nos permite, entre outras coisas, medir o nível de expressão de milhares de genes simultaneamente e determinar parcialmente a topologia da rede de regulação gênica [14]. Infelizmente, nenhuma técnica até o momento nos permite conhecer a regra que rege o processo de regulação. Esse é um dos pontos em que modelos computacionais podem ser úteis. A elaboração de modelos nos quais sejam sugeridas regras para a regulação e a posterior comparação com os dados experimentais pode validar uma determinada regra. De fato, modelos assim têm sido propostos e com resultados animadores [15]. Entre esses resultados, o que de mais perto nos toca é o fato de que regras simples do tipo limiar parecem descrever bem o mecanismo de regulação. Motivados por isso, neste capítulo iremos propor uma regra do tipo limiar e estudaremos o tipo de comportamento que ela gera. Esse estudo da dinâmica de uma célula nos permitirá analisar, no próximo capítulo, o fenômeno da diferenciação celular.

5.1 A Escolha da Regra

No capítulo anterior, construímos a rede de regulação gênica, ou seja, definimos como os genes estão conectados uns com os outros e se essa conexão é do tipo ativadora ou inibidora. Entretanto, ainda não dissemos de que forma os genes que estão conectados interagem uns com os outros. Por exemplo, se um gene regulado por dois outros genes recebe de um deles um estímulo ativador e de outro um estímulo inibidor, como ele deve se comportar? Oferecemos nessa seção uma proposta simples para essa questão.

No nosso modelo, os genes possuem somente dois estados: ativos ($\sigma = 1$) ou inativos ($\sigma = 0$). Se um gene está ativo, então ele está sendo transcrito e a proteína que ele codifica será sintetizada na célula. Caso esteja inativo, não haverá transcrição. Propomos que uma vez definido o estado de todos os genes em um tempo t , nós saberemos o estado de cada um dos genes em um tempo $t + 1$ por meio da seguinte regra de evolução ou função de atualização:

$$\sigma_i(t + 1) = \text{Sgn}(J_{ii}\sigma_i(t) + \sum_{j=1}^{k_i} J_{j(i)}\sigma_{j(i)}(t) - \theta) \quad (5.1)$$

$$\text{Sgn}(x) = \begin{cases} 1 & \text{se } x > 0 \\ 0 & \text{se } x \leq 0 \end{cases} \quad (5.2)$$

A constante de acoplamento $J_{j(i)}$ nos dá a informação de como o gene j regula o gene i , que pode ser de forma a ativar ou inibir a transcrição. Em nosso modelo, a constante de acoplamento têm o valor $+1$ ou -1 caso o gene j esteja ativando ou inibindo a transcrição do gene i , respectivamente. O valor das constantes de acoplamento é escolhido aleatoriamente com probabilidade $P1$ de ser inibidora e $1 - P1$ de ser ativadora. Dessa forma, o que a regra nos diz é que para decidir o seu estado em um tempo $t + 1$ um gene soma todos os estímulos devido aos genes que o regulam e a si mesmo ($J_{ii}\sigma_i(t) + \sum_{j=1}^k J_{j(i)}\sigma_{j(i)}(t)$) em um tempo t , se essa soma for maior que um determinado limiar de ativação θ esse gene estará ativo em $t + 1$, caso contrário estará inativo. Observe que mesmo se um gene j faz parte dos reguladores de um gene i , ele

somente estará exercendo efetivamente a regulação se estiver ativo. O motivo disso é que se um gene regulador está inativo não existe a síntese do fator de transcrição que ele codifica. Logo, esse fator não atuará na regulação de outros genes.

Cada vez que aplicamos a regra (5.1) sobre uma rede, dizemos que estamos evoluindo o sistema. O nosso sistema é evoluído a partir de configurações iniciais escolhidas aleatoriamente. Em todos os resultados que serão apresentados a seguir, nós consideramos o limiar de ativação $\theta = 0$. Escolhemos esse valor porque é o valor padrão utilizado em regras do tipo limiar [15] e também porque parece ser o que mais se aproxima do correspondente em redes reais [16]. Os resultados obtidos com outros limiares de ativação são qualitativamente equivalentes àqueles obtidos com o limiar $\theta = 0$, e por esse motivo não serão apresentados.

5.2 Atrator, Período e Transiente

Suponhamos uma rede que possui N genes. Como cada um dos genes somente pode estar ativo ou inativo, existem um total de 2^N possíveis configurações diferentes. Cada configuração representa um padrão possível de expressão gênica. Começamos de uma configuração inicial arbitrária e apliquemos sucessivamente a regra de atualização (5.1) sobre ela. O que acontecerá? Cada vez que aplicarmos a regra de atualização o sistema “pulará” para uma nova configuração. Mas isso não acontece indefinidamente, nem poderia uma vez que o número de configurações é finito. O que observamos é que depois de passar por um determinado número de configurações o sistema retorna a uma configuração já visitada, e a partir daí começa a repetir toda uma seqüência de configurações ciclicamente. Quando isso acontece, dizemos que o sistema está percorrendo o seu atrator. Um atrator é o que o próprio nome diz: uma seqüência de configurações que atrai outras configurações e uma vez que se está dentro dele não se pode mais sair.

O espaço de configurações está dividido entre vários atratores, formando estruturas similares a da figura 5.1. Obviamente a estrutura do espaço de configurações

é distinta para cada rede diferente que escolhermos, diferindo quanto ao número de atratores, ao número de configurações que cada atrator possui, bem como ao número de configurações que cada atrator atrai para si. Por isso, estaremos interessados nos valores médios dessas quantidades sobre um *ensemble* de várias redes.

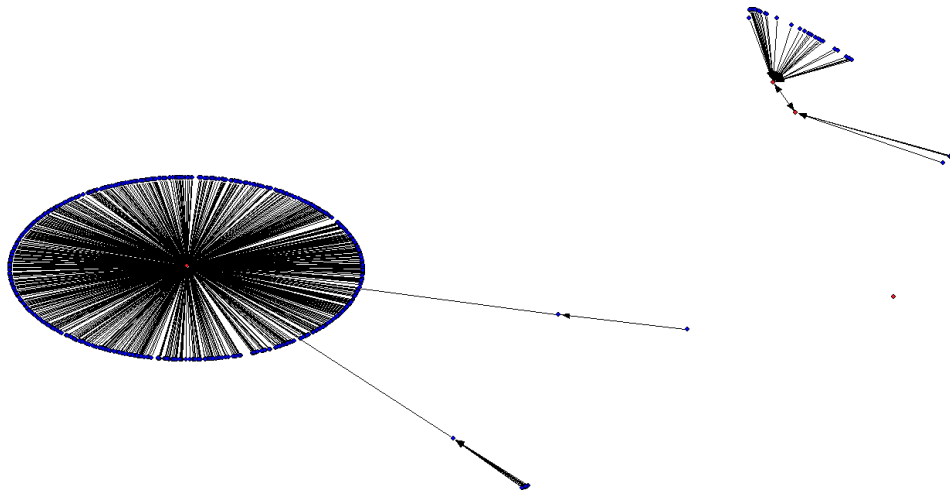


Figura 5.1: Exemplo de estrutura do espaço de configurações. Cada ponto representa uma configuração e as setas indicam para onde cada uma das configurações converge. Na figura notamos 3 atratores, sendo 2 deles pontos fixos e um deles um ciclo de período 2. Figura gerada para uma rede com 10 nós, $P1=0.1, \gamma = 1.1$, correspondente ao regime congelado (será definido posteriormente).

Chamamos de período do atrator ao número de configurações que o atrator possui. Chamamos de transiente ao número de configurações que o sistema percorre partindo de uma configuração inicial antes de entrar no seu atrator. Ao conjunto de configurações que são atraídas por um dado atrator mais as configurações pertencentes aquele atrator, chamamos de bacia de atração do dito atrator.

A grande importância dos atratores está na associação que se faz entre eles e os ciclos de funcionamento das células. Espera-se que as configurações que fazem parte do atrator, bem como a ordem em que elas aparecem, representem os vários estados de expressão gênica ao longo do ciclo celular. Essa associação nos permite ir mais além. Podemos dizer que células que operam em atratores diferentes representam

tipos celulares diferentes, já que possuem diferentes padrões de expressão gênica. Essa será uma das hipóteses fundamentais de nosso trabalho. Todas essas idéias são devidas a Stuart Kauffman [17].

5.3 Os Três Regimes de Comportamento

O material genético dos organismos está sujeito a alterações (mutações) que ocorrem tanto devido à presença de fatores externos, como substâncias mutagênicas (alcatrão, nicotina, aflatoxinas, etc.) e radiação; bem como devido a erros inerentes ao processo de replicação do DNA [5]. Apesar disso, as células são robustas contra mutações. Como observado por Visser *et al.*[18], “robustez é a invariância do fenótipo em face de perturbações”, ou seja, embora haja alterações no material genético dos organismos, na maior parte das vezes eles continuam funcionando de maneira adequada. É um fato que parte dessa robustez é devido a sofisticados mecanismos a nível genético, tais como redundância genética e “*buffering epistasis*”¹. Mas existem evidências sugerindo que a robustez genética também pode estar associada com propriedades dinâmicas e estruturais da rede de interação genética [18]. Sendo assim, pretendemos verificar se o nosso modelo incorpora essa tolerância ao erro. Com esta finalidade, utilizaremos o procedimento de evoluir uma rede a partir de uma configuração inicial aleatória qualquer até que ela entre em alguma configuração que pertença a um dos seus atratores. Quando isso acontecer, faremos uma cópia dessa configuração e inseriremos nela uma pequena modificação, algo como alguns poucos genes escolhidos ao acaso que terão o seu estado trocado de ativo para inativo ou vice-versa. Depois disso feito, evoluiremos a rede partindo dessas duas configurações (a original e a cópia) e mediremos a cada passo de evolução o quanto cada uma das configurações para a qual o sistema evolui difere da outra. Essa medida de o quanto duas configurações

¹*buffering epistasis* refere-se ao fenômeno no qual um gene i é regulado por dois outros genes da mesma maneira e a inativação de um desses genes não ocasionará nenhuma mudança perceptível na regulação de i .

diferem entre si é chamada distância de Hamming e é dada pela expressão a seguir:

$$\Psi(t) = \frac{1}{N} \sum_{i=1}^N |\sigma_i(t) - \sigma_i^a(t)| \quad (5.3)$$

Conforme vemos pela fórmula 5.3, a distância de Hamming nada mais é do que o número de genes nos quais as duas configurações $\{\sigma(t)\}$ e $\{\sigma^a(t)\}$ diferem dividido pelo número de genes que a rede possui (N).

Vamos supor que as duas configurações (a original $\{\sigma(t)\}$ e a cópia $\{\sigma^a(t)\}$) diferem inicialmente em K genes. Portanto, a distância de Hamming inicial ($\sigma(0)$) será K/N . Se depois de um tempo t suficientemente grande a distância de Hamming tender para zero, dizemos que a rede opera no regime congelado; se for igual a K/N dizemos que opera no regime marginal e se for maior que K/N dizemos que opera no regime caótico. No regime congelado, um dano introduzido quase certamente será corrigido pela dinâmica da rede, tendendo na maior parte das vezes para zero desde que N seja suficientemente grande. Na fase marginal, o dano introduzido não é corrigido, mas também não se espalha para outros genes da rede. Na fase caótica o dano se espalha pela rede atingindo outros sítios e podendo fazer as duas configurações diferirem enormemente, mesmo que a diferença inicial seja mínima.

Em qual desses regimes as redes biológicas operam? A resposta clássica a isso tem sido o regime marginal. Na expressão de Kauffman, a vida acontece na “borda do caos”. O regime caótico definitivamente não é uma opção viável, uma vez que uma pequena modificação pode causar mudanças drásticas, ou seja, não temos robustez (existe pelo menos uma evidência direta, baseada no estudo de uma rede real, de que as células não podem ser caóticas [18]). O regime congelado, por sua vez, é compatível com a resistência a erros das redes reais, mas parece ser incompatível com a evolução natural, já que não aceita nenhum tipo de mutação. Portanto, sobra somente o regime marginal, que reúne uma resistência a erros e uma mutabilidade que permite a evolução natural.

Vamos estudar como os três regimes aparecem em nosso modelo.

Na figura 5.2, vemos o valor da razão entre a distância de Hamming para tempos

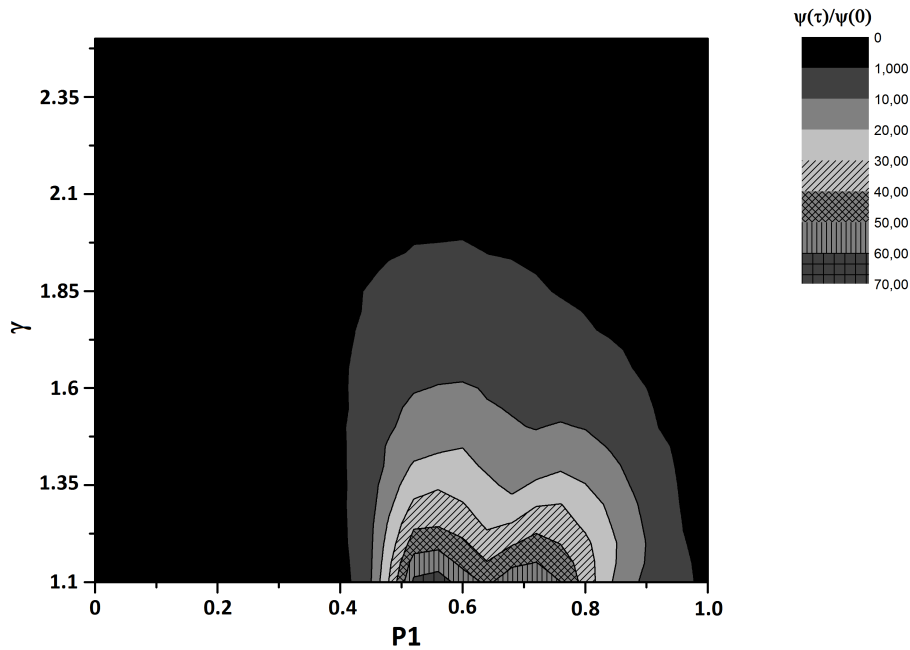


Figura 5.2: Espaço dos danos com limiar de ativação $\theta = 0$. Média de *ensemble* (8000 redes) sobre redes com $N = 1000$ genes reguladores ($perc = 1.0$). O dano inicial é de apenas um gene.

suficientemente grandes e a distância de Hamming inicial em função das variáveis $P1$ e o expoente γ da distribuição livre de escala. A distância de Hamming inicial é a pequena diferença que se insere nas duas configurações iniciais que são evoluídas. Com esse gráfico podemos identificar os três regimes. O regime congelado corresponde a toda faixa para a qual $\Psi(t)/\Psi(0)$ é menor que 1, que corresponde a toda área em negro do gráfico. O regime marginal é a linha onde $\Psi(t)/\Psi(0)$ é exatamente igual a 1. O regime caótico corresponde a toda faixa onde $\Psi(t)/\Psi(0)$ é maior que 1. Na figura 5.2, o regime caótico está dividido em vários níveis com intuito de mostrar como $\Psi(t)/\Psi(0)$ varia nesse regime. Nele pode ser observado que quanto menor é o expoente mais caótico o sistema tende a ser.

É notável a alta sensibilidade a danos iniciais no regime caótico. Observe que para algumas regiões o dano final pode chegar a ser 70 vezes maior do que o dano inicial

introduzido no sistema. Seria impossível uma célula funcionar nesse regime, uma vez que um pequeno dano levaria a célula para outro padrão de expressão totalmente distinto.

O que podemos dizer sobre a previsão do nosso modelo e o observado nos organismos? As evidências experimentais apontam que estamos no caminho correto. A faixa de valores do expoente γ no nosso modelo para a qual existe uma região marginal é justamente aquela na qual encontramos os expoentes das redes reais de regulação. Diferentemente da maior parte das outras redes livre de escala, que tendem a apresentar expoentes maiores que 2 [19], as redes de regulação parecem ser um caso especial, já que tendem a apresentar expoentes menores do que 2 [11]. Infelizmente, não podemos verificar a acurácia de nosso modelo quanto ao valor de $P1$, já que os dados na literatura são muito escassos.

Podemos inferir do nosso modelo que regras do tipo limiar são mais apropriadas para o estudo da regulação do que as intensivamente utilizadas funções Booleanas. Em um trabalho onde foi utilizada uma rede equivalente a nossa, Aldana escolheu as funções Booleanas como regra de atualização [19]. Como resultado, a faixa de valores para o expoente onde se encontra o regime marginal é entre 2 e 2.5. Além disso, o trabalho prevê que para expoentes menores que 2 as redes operam no regime caótico, independente de qualquer outro parâmetro do modelo. Ora, sabemos que é justamente nessa faixa que as redes reais se encontram e elas não são caóticas.

5.4 A Dinâmica da Rede é Independente do Grupo dos Genes Estruturais

Os genes estruturais são de importância fundamental para vida. Na verdade, são um pré-requisito mais fundamental do que os genes reguladores, já que temos organismos que funcionam quase sem nenhuma regulação. Mas será que a existência dos genes estruturais tem algum efeito na dinâmica total do processo de regulação? O nosso modelo indica que não. Conforme observamos, as quantidades fundamentais

que determinam a dinâmica do sistema: tamanho do período dos atratores, tamanho dos transientes, número de atratores, etc., são independentes da existência do grupo dos genes estruturais (figura 5.3). Esse resultado nos fornece várias vantagens de ordem computacional como, por exemplo, a diminuição do número de parâmetros do modelo e uma grande diminuição do número de genes que nossa rede precisa ter para equiparar-se com redes reais. De fato, se tivéssemos que considerar os dois conjuntos de genes para o *S. cerevisiae*, por exemplo, teríamos um total de 491 genes, enquanto considerando somente o grupo dos genes reguladores temos 124 genes [20].

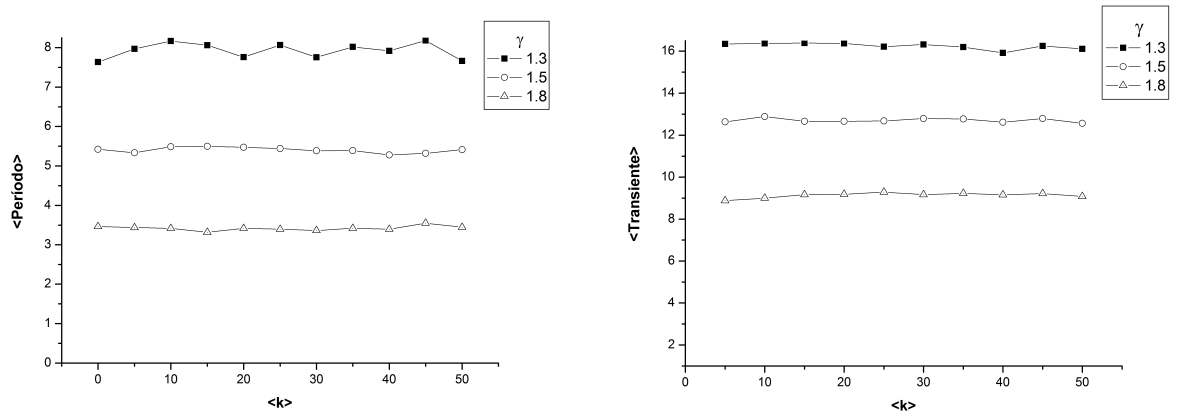


Figura 5.3: Tamanho médio do (a) período e do (b) transiente dos atratores em função de $\langle k \rangle$. Observe que as duas grandezas são independentes da maneira como se dá a interação do grupo dos genes estruturais com os reguladores e são mesmo independentes da existência dessa ligação ($\langle k \rangle = 0$). Sendo assim, a existência do grupo dos genes estruturais não altera a dinâmica do sistema.

5.5 O Tamanho dos Períodos e Transientes

Observemos a figura 5.4, na qual é apresentado o tamanho médio dos períodos e dos transientes em função do parâmetro $P1$ para o expoente $\gamma = 1.3$. Se compararmos essa figura com a figura 5.2, podemos verificar que o regime caótico, em sua maior

parte, é caracterizado por grandes períodos e grandes transientes. O regime marginal, por sua vez, possui em média períodos e transientes maiores do que o regime congelado, mas ainda bem pequenos se comparados com o regime caótico.

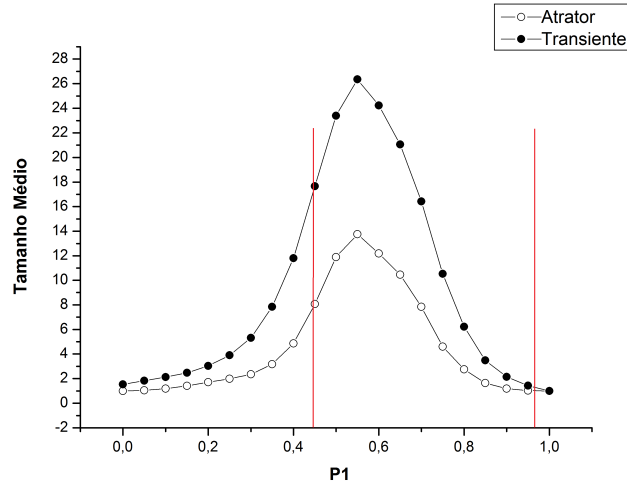


Figura 5.4: Tamanho médio do transiente e do período dos atratores em função de $P1$. A linha vermelha indica o regime marginal. Rede com $N = 100$ e $\gamma = 1.3$.

Na figura 5.5, estudamos como o valor médio do período varia com o número de genes da rede nos três regimes. Observamos que no regime caótico o período médio dos atratores aumenta exponencialmente, no regime marginal linearmente e no regime congelado ele tende a saturação. O tamanho do período dos atratores está relacionado de alguma forma à duração do ciclo celular, provavelmente proporcional ao tempo de duração do ciclo celular. Vamos fazer um cálculo grosseiro: suponhamos que uma célula percorra o seu atrator somente uma vez a cada ciclo celular, se levarmos em conta que em células eucariotas os genes demoram cerca de 10 minutos para se tornarem ativos e que o ciclo celular tem um tempo de duração típico de 3700 minutos, limitamos o número de configurações que um atrator pode ter em $370[21]$. O que queremos concluir com tudo isso é a impossibilidade de uma célula operar no regime caótico. Como o período dos atratores aumenta exponencialmente, mesmo para pequenas redes, esse número é rapidamente excedido.

Observamos da figura 5.2, que existem duas zonas marginais para cada expoente

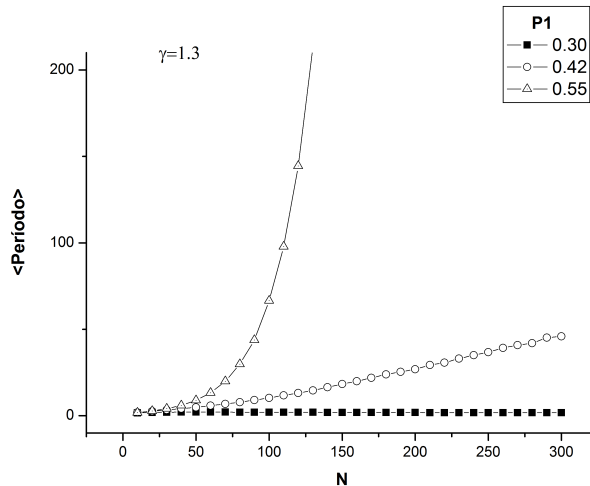


Figura 5.5: Tamanho médio do período nos três regimes de comportamento do modelo.

menor que $\gamma = 2.0$. Comparando as figuras 5.2 e 5.4, observamos que a primeira figura não possui simetria, enquanto a segunda figura possui simetria em torno de $P1 = 0.55$. Devido a essa diferença, o valor médio do período e do transiente difere bastante se uma rede opera, por exemplo, na zona marginal à esquerda de $P1 = 0.55$ ou na zona marginal à direita. Essa diferença nos motiva a perguntar se as duas zonas marginais são aceitáveis do ponto de vista biológico. Parece-nos que não. O motivo é que a maior parte das redes na zona marginal à direita tende ao ponto fixo em que todos os genes permanecem sempre inativos. Isso é inaceitável biologicamente, posto que se todos os genes estiverem inativos não haverá produção de nenhum tipo de fator de transcrição ou proteína. Se a zona marginal à esquerda levasse para o ponto fixo onde todos os genes estão ativos ela também não seria aceitável, visto que os dados biológicos sobre a expressão dos genes não apontam nesse sentido [22]. Mas, como podemos ver na figura 5.6, não é isso o que acontece. A zona marginal à esquerda contém vários genes inativos e ativos, o que corresponde ao caso biológico. A título de informação, o percentual de genes expressos entre os diferentes tipos celulares humanos se encontra entre 42% e 57%.

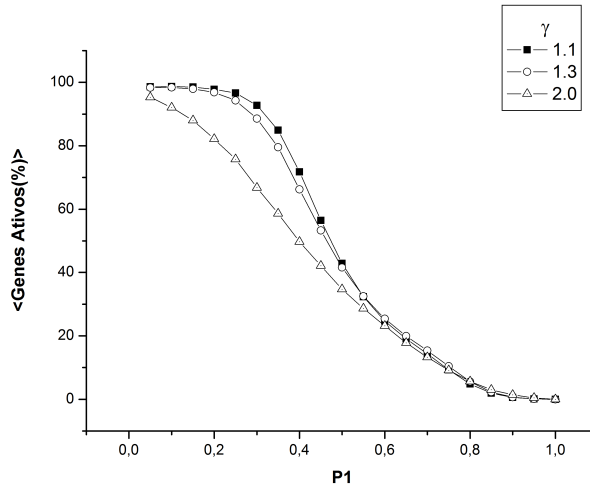


Figura 5.6: Percentual médio de genes ativos em função de $P1$. Rede com $N = 100$ genes.

5.6 O Número de Atratores

Contar o número de atratores é uma tarefa que se torna impraticável a partir de certo ponto. A grande questão é que para contar os atratores nós precisamos determinar para qual atrator cada configuração inicial converge, uma vez que uma rede com N nós possui 2^N configurações iniciais essa tarefa se torna computacionalmente impossível mesmo para pequenos valores de N . Dessa forma, a abordagem que utilizamos foi a de escolher um número fixo $C.I$ de configurações iniciais para cada rede e contar para quantos atratores elas convergem. Parece-nos que com essa abordagem poderemos ao menos dizer em qual dos três regimes de comportamento existem mais atratores. Na figura 5.7, apresentamos como o número médio de atratores varia nos três regimes de comportamento utilizando a abordagem citada.

Observamos que o número de atratores é máximo no regime caótico. De fato, isso explica o porquê desse regime ser tão sensível a mutações: quanto maior o número de atratores, maiores são as chances de que uma mutação conduza um sistema de um atrator para outro. O número de atratores no regime congelado diminui com o tamanho da rede, sendo que os principais atratores nesse regime são os dois pontos

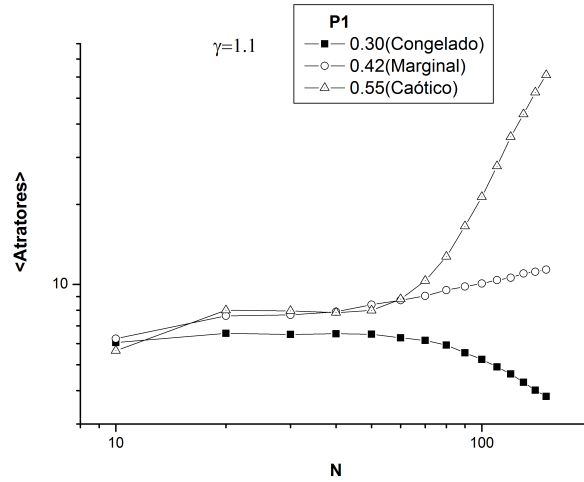


Figura 5.7: Número médio de atratores nos três regimes. Cada ponto é uma média sobre 14.000 redes com 100 configurações iniciais para cada uma.

fixos onde todos os genes estão ativos ($P1$ indo para zero) ou inativos ($P1$ indo para um). O regime marginal mantém as suas características de região intermediária: não tem tantos atratores como o regime caótico e nem tão poucos quanto o regime congelado.

Uma informação importante que gostaríamos de obter é se o número de atratores varia com N segundo alguma função conhecida. Utilizando a abordagem que propomos poder-se-ia dizer que no regime marginal, por exemplo, o número de atratores aumenta com uma lei de potência. Entretanto, uma rápida observação na figura 5.8 nos mostra que a nossa abordagem não nos permite obter essa informação. Na figura 5.8, nós comparamos diversas curvas geradas a partir de um número fixo de diferentes configurações iniciais $C.I$ com a solução “exata”, na qual analisamos todas as configurações. Como se pode observar, o número de atratores varia segundo funções diferentes, por assim dizer, conforme variamos o $C.I$.

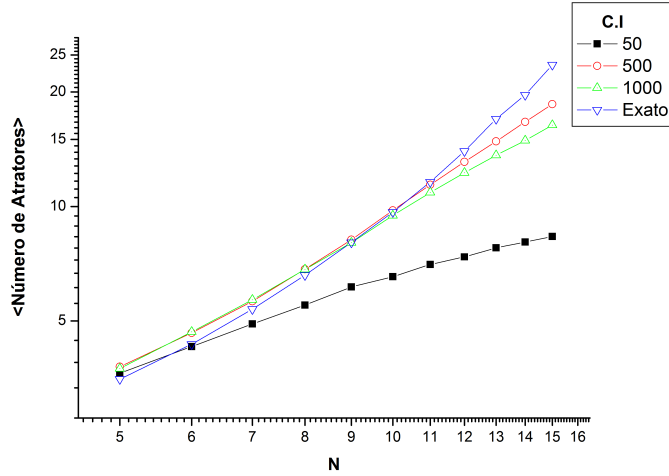


Figura 5.8: Número médio de atratores nos três regimes em função do número de configurações iniciais ($C.I.$). Cada ponto é uma média sobre 10.000 redes com $C.I.$ configurações iniciais para cada uma.

5.7 A Assimetria na Figura 5.2

A assimetria do nosso espaço de danos (Figura 5.2) é resultado do diferente *modus operandi* de genes inativos e ativos: genes inativos não exercem regulação e genes ativos sim. Com o intuito de verificar isso, podemos propor uma regra que trate genes inativos e ativos da mesma forma. Seja, por exemplo, a seguinte regra:

$$\sigma_i(t+1) = Sgn(J_{ii}\sigma_i(t) + \sum_{j=1}^{k_i} J_{j(i)}(2\sigma_{j(i)}(t) - 1) - \theta) \quad (5.4)$$

$$Sgn(x) = \begin{cases} 1 & \text{se } x > 0 \\ 0 & \text{se } x \leq 0 \end{cases} \quad (5.5)$$

É fácil ver que nessa regra um gene inativo exerce regulação. Ao fazermos $\sigma_{j(i)}(t) = 0$ obteremos que o estímulo regulatório do gene j sobre o gene i é dado por $-J_{j(i)}$, e não 0 como na regra que utilizamos. De fato, essa regra admite mesmo uma interpretação biológica plausível. Se o gene j está inativo e ele regula o gene i positivamente (negativamente), ou seja, $J_{j(i)} = 1$ ($J_{j(i)} = -1$), então o estímulo regulatório enviado será negativo (positivo), significando que ele tende a inibir (ativar) a regu-

lação no sentido de não estar produzindo o fator de transcrição que lhe é associado, ou seja, uma inibição (ativação) por omissão. A interpretação quando o gene j está ativo é análoga.

Na figura 5.9 geramos o espaço de danos associado à regra citada. Como podemos ver a figura 5.9 apresenta a simetria ausente na figura 5.2.

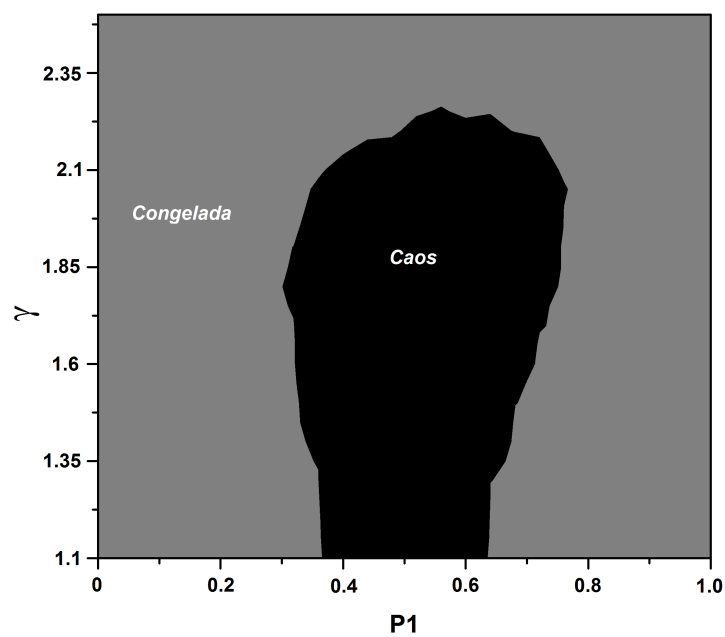


Figura 5.9: Espaço de danos para a regra de atualização. Observe a simetria presente na figura. Limiar de ativação $\theta = 0$. Média de *ensemble* (8000 redes) sobre redes com $N = 500$ genes reguladores ($perc = 1.0$). O dano inicial é de apenas um gene.

Capítulo 6

Um Modelo para a Diferenciação Celular

Os organismos multicelulares são formados por diferentes tipos de células, que diferem entre si tanto estrutural quanto funcionalmente. Apesar da estonteante complexidade de organismos como nós, por exemplo, sabemos que tudo começa de uma forma simples, a partir de uma única célula que se divide rapidamente. O aparecimento de novos tipos celulares, processo chamado de diferenciação celular, acontece durante esse processo de desenvolvimento, que é bastante complicado e não é completamente entendido.

Neste capítulo, propomos um modelo para o aparecimento de novos tipos celulares ao longo dos estágios iniciais do desenvolvimento embrionário. Em nosso modelo, o aparecimento de novos tipos celulares é consequência de uma característica que julgamos imprescindível para o fenômeno da diferenciação celular: a existência de comunicação entre diferentes células. Como veremos, segundo o nosso modelo, esse único requisito é o bastante para gerar, em vários aspectos, diferenciação similar a observada em organismos reais.

6.1 O Modelo

No capítulo 5, estudamos a dinâmica de uma rede de regulação gênica. Como já foi dito, o que caracteriza uma célula como sendo de um tipo ou outro, é justamente os seu padrão de expressão gênica, que é determinado pela dinâmica de sua rede de regulação. Dessa forma, nada mais natural do que utilizarmos o modelo já desenvolvido no capítulo 5 e ampliá-lo, permitindo agora que várias redes de regulação, que correspondem às nossas células, interajam entre si.

Cada célula em nosso modelo divide-se sobre uma rede quadrada bidimensional depois de transcorrido um tempo t desde a sua última divisão. Em analogia com o caso biológico, inicialmente temos uma única célula. Quando uma célula se divide (célula mãe), ela gera outra célula idêntica (célula filha), ou seja, a nova célula possui a mesma rede, as mesmas constantes de acoplamento e está no mesmo estado da célula que lhe deu origem. A posição que a célula filha irá ocupar na rede quadrada é escolhida aleatoriamente entre os sítios da vizinhança de Moore (1^o e 2^o vizinhos) da célula mãe (veja a figura 6.1). Caso essa posição esteja ocupada, ela deslocará as outras células ao longo dessa direção.

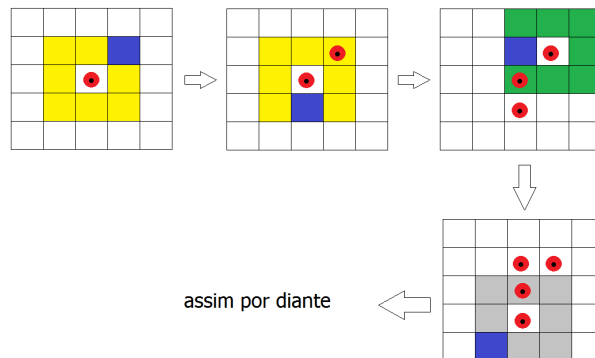


Figura 6.1: Desenho esquemático da forma como as células se dividem. A vizinhança de Moore corresponde aos sítios coloridos em volta de uma célula (os seus 8 vizinhos). O sítio de cor azul escuro pertencente à vizinhança de Moore de uma dada célula representa a posição que sua célula filha irá ocupar.

Nós inserimos a interação celular em nosso modelo mediante a criação de novos

acoplamentos entre genes pertencentes a células distintas. Quando uma célula se divide, ela e a célula filha escolhem independentemente e aleatoriamente uma célula das suas vizinhanças de Moore. Um dos genes reguladores dessas células escolhidas, também selecionado ao acaso, irá regular um gene pertencente à célula que o escolheu (a célula mãe ou a filha). O gene que será regulado na célula mãe ou na célula filha é selecionado aleatoriamente entre aqueles que fazem parte do grupo que é regulado pelo gene regulador. Sendo assim, a função de atualização 5.1 para cada gene de uma célula irá, conforme o número de células aumenta, recebendo mais e mais acoplamentos devido à regulação por parte de genes pertencentes a outras células. A nova função de atualização será dada agora por:

$$\sigma_i^\alpha(t+1) = \text{Sgn}(J_{ii}\sigma_i^\alpha(t) + \sum_{i=1}^{k_i} J_{j(i)}\sigma_{j(i)}^\alpha(t) + \sum_{\text{Aopl. Intercel. } j} J_{j(i)}\sigma_{j(i)}^\beta(t) - \theta) \quad (6.1)$$

onde o segundo somatório envolve todos os acoplamentos que o gene i da célula α recebe de genes de outras células.

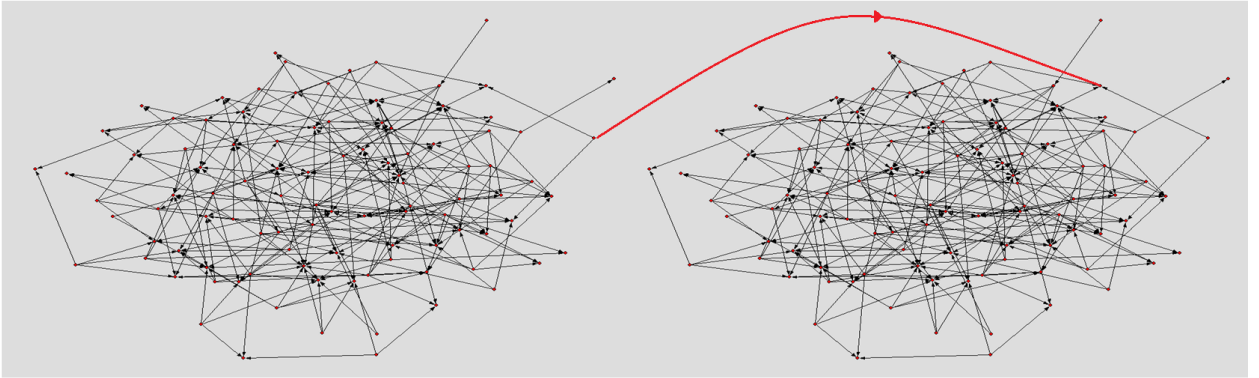


Figura 6.2: A interação intercelular é feita através da regulação entre genes pertencentes a células distintas.

Em resumo, nosso modelo assume que as interações intercelulares são principalmente entre primeiros e segundos vizinhos, portanto locais. De um ponto de vista biológico, estamos assumindo que as células interagem principalmente via a troca de sinais por meio das junções *gap* formadas entre elas. Devido aos deslocamentos celulares provocados pelas divisões mitóticas, acontece também de células interagirem

com outras células que não pertencem mais à sua vizinhança de Moore. De fato, esse mecanismo de sinalização à distância acontece em organismos reais, já que certos sinais indutores podem se difundir no meio e ter um alcance mais amplo de ação. Já a interação célula-célula é desordenada, pois o gene regulador é escolhido ao acaso em cada interação, e a mais fraca possível porque envolve apenas um par de genes por vez. Como demonstraremos a seguir, mesmo esse acoplamento mínimo e aleatório célula-célula é capaz de gerar diferenciação celular.

6.2 Resultados

No capítulo anterior, concluímos que o regime marginal parece ser o mais adequado para o funcionamento de uma célula. Como veremos a seguir, os resultados obtidos quando consideramos a interação de várias células corroboram esse ponto de vista. Dos três regimes, o regime marginal é aquele que apresenta um comportamento mais próximo, no que diz respeito ao aparecimento de novos tipos celulares, daquele observado nos organismos.

Na figura 6.3, podemos observar como é o aparecimento de novos tipos celulares com o aumento do número de células em nosso modelo. O regime congelado não apresenta aumento no número de tipos celulares e, portanto, não é aceitável para a descrição de organismos multicelulares. O regime caótico, por sua vez, apresenta o aparecimento de novos tipos celulares em demasia e com oscilações complexas no número de tipos celulares que não são observadas nos organismos reais. Por outro lado, é notável a semelhança entre o comportamento no regime marginal e aquele dos organismos reais *T. torumatics* e *C. elegans*, que pode ser visto no anexo da figura 6.3.

Com intuito de apresentar informações que ficam ocultas na figura 6.3, construímos a linhagem celular típica dos três regimes de comportamento. A linhagem celular pode ser entendida como a árvore genealógica do desenvolvimento de um organismo. Nela, encontramos informações como que célula gerou qual e quais os tipos celulares

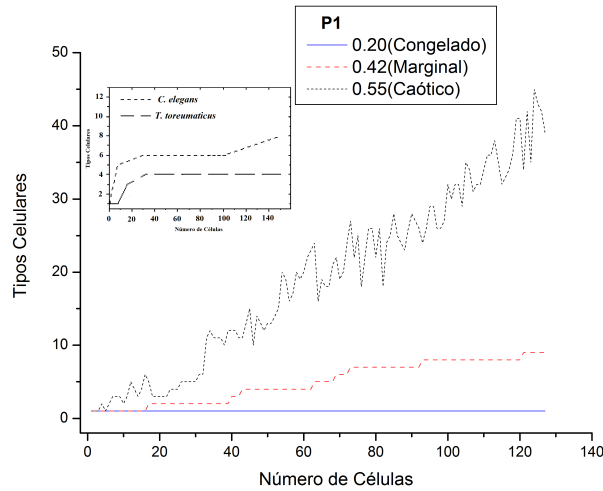


Figura 6.3: Número de tipos celulares em função do número de células. Rede com $N = 100$ genes e expoente $\gamma = 1.1$. No anexo, observa-se o aparecimento de novos tipos celulares em organismos reais. Anexo retirado e modificado de [26].

gerados. Na figura 6.4, vemos a linhagem celular típica do regime congelado. Como já havíamos observado da figura 6.3, o regime congelado não apresenta aumento do número de tipos celulares, ou seja, não existe diferenciação celular.

Na figura 6.5, vemos a linhagem celular típica para uma célula operando no regime caótico. O mais notável na figura 6.5, e que inviabiliza o regime caótico como opção biológica, é que não existe o aparecimento de tipos celulares permanentes. De fato, no regime caótico realmente aparecem novos tipos celulares, mas eles continuam sempre mudando ao longo do desenvolvimento, o que está em flagrante oposição com o caso biológico, aonde as células vão se tornando cada vez mais determinadas para um destino celular específico.

O regime marginal, como podemos ver comparando a figura 6.6 e a figura 6.7, é aquele que mais se assemelha ao caso biológico. Além de produzir novos tipos celulares, esses tipos celulares também alcançam efetivamente um estado no qual estão completamente determinados e, portanto, não mais se modificam.

A morfogênese é a etapa no desenvolvimento em que se dá a formação de tecidos e órgãos. Alguns dos mecanismos conhecidos mais importantes para a morfogênese

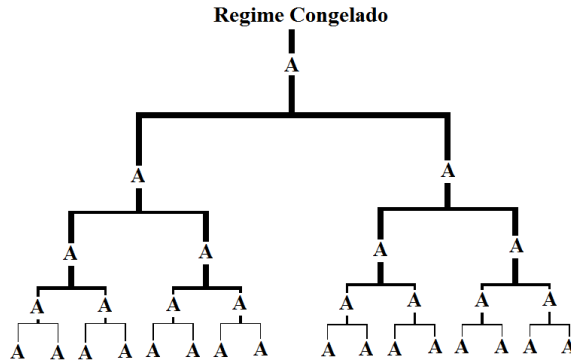


Figura 6.4: Linhagem celular típica gerada a partir de uma célula operando no regime congelado. A rede contém $N = 100$ genes, $\gamma = 1.1$ e $P1 = 0.20$.

são: a adesão celular diferenciada, a informação posicional, migração celular e a apoptose. Nenhum desses mecanismos está presente em nosso modelo. Por essa razão, ele se enquadra bem para explicar o início do desenvolvimento, quando esses mecanismos não são muito atuantes. Apesar disso, mostramos na figura 6.8, os padrões espaciais típicos gerados pelos três regimes de comportamento. Como podemos observar, os padrões gerados pelo regime marginal e caótico são aleatórios, não produzindo a formação de domínios extensos contidos pelo mesmo tipo celular, que poderiam ser associados à formação dos diferentes tecidos. Dessa forma, o nosso modelo indica que, na ausência dos mecanismos que são comumente atribuídos como responsáveis pela morfogênese, ela não ocorre. Entretanto, é necessário ressaltar que H. S. Silva e M. L. Martins [24] em trabalho similar ao nosso, chegaram a uma conclusão diferente. De fato, mesmo na ausência dos mecanismos da morfogênese já citados, o modelo que eles propõem, interessantemente, apresenta formação de tecidos no regime marginal. O mais notável disso tudo, é que o nosso modelo praticamente difere do modelo citado somente na escolha da rede de regulação gênica, o que sugere que a topologia da rede pode ter um papel importante na determinação da morfogênese.

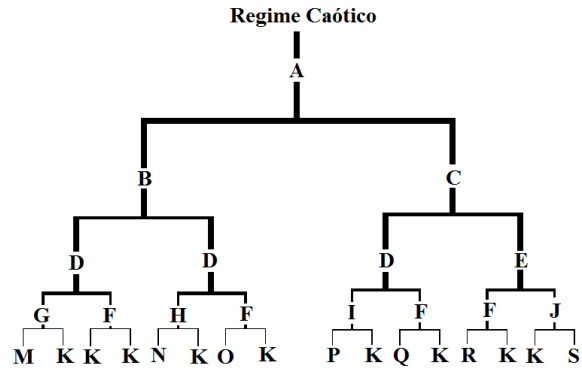


Figura 6.5: Linhagem celular típica gerada a partir de uma célula operando no regime caótico. A rede contém $N = 100$ genes, $\gamma = 1.1$ e $P1 = 0.55$.

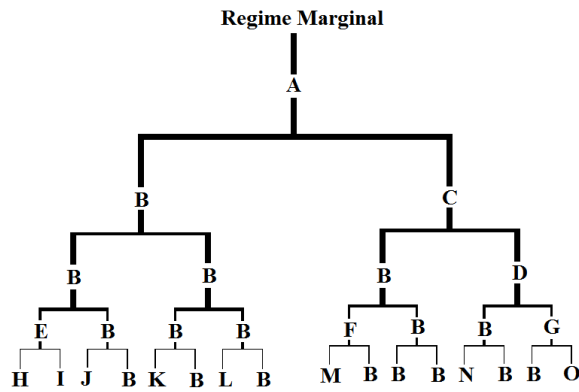


Figura 6.6: Linhagem celular típica gerada a partir de uma célula operando no regime marginal. A rede contém $N = 100$ genes, $\gamma = 1.1$ e $P1 = 0.42$.

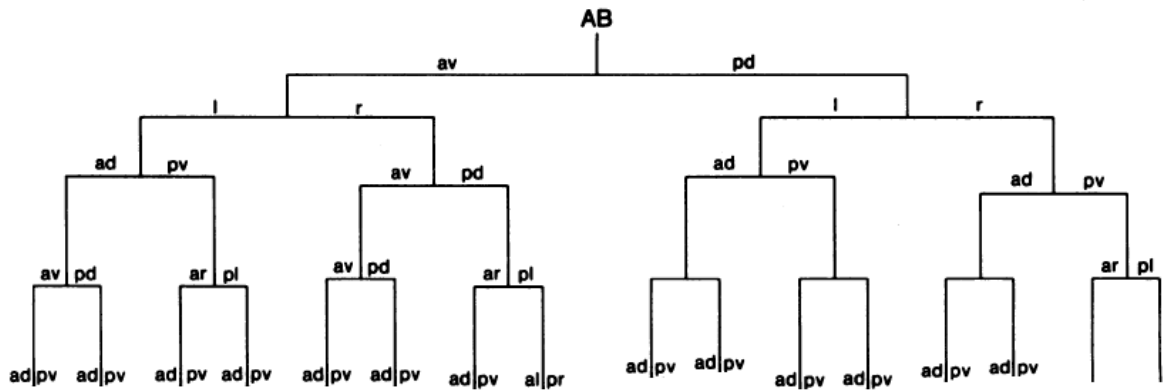


Figura 6.7: Uma parte da árvore de linhagem celular do nematódeo *C. elegans*. Observe a conservação através de diferentes ramos de tipos celulares como av e pd, por exemplo. Retirado de [23].

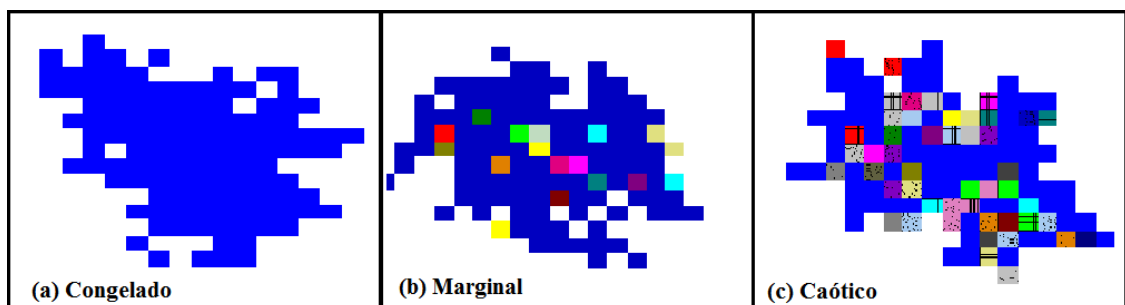


Figura 6.8: Padrões típicos gerados pelos três regimes. No regime congelado (a) temos 1 tipo celular; (b) 15 tipos e no regime caótico (c) 50 tipos. Células $N = 100$ genes e os valores de γ e $P1$ como os usados na figura 6.2 foram considerados.

Capítulo 7

Conclusões

DECIFRAR o genoma não é o bastante para compreender todas as características de um organismo. Ao que parece, é necessário também estabelecer como os genes interagem entre si, ou seja, a topologia da rede de regulação gênica, e também, como se dá essa dinâmica de regulação. Questões de importância singular como a explicação do funcionamento, da estrutura e da complexidade dos organismos, devem emergir como resultado dessa compreensão. Segundo nosso modelo, redes de regulação gênica podem operar em três tipos de regime de comportamento, a saber: congelado, marginal e caótico. Entretanto, somente redes operando no regime marginal reúnem as características necessárias às redes reais de regulação, como estabilidade a danos e evolução através de mutações não-deletérias. Nosso modelo ainda revela que a faixa de valores do expoente da distribuição de conectividades da rede para a qual existe o regime marginal, é justamente a faixa na qual as redes reais de regulação se situam.

O nosso modelo para a diferenciação celular sugere que o aparecimento de diferentes tipos celulares ao longo da embriogênese é resultado de uma mudança de atratores das redes de regulação das células como consequência de interações entre as mesmas. A diferenciação celular gerada a partir de células que operam no regime marginal está de acordo com características observadas no desenvolvimento de or-

ganismos reais como: a taxa de aparecimento de novos tipos celular ao longo do desenvolvimento e a existência de células que se tornam completamente determinadas a um destino celular específico. Em concordância com o novo paradigma que sugere a existência de características fenotípicas codificadas na topologia da rede de regulação gênica [2], o nosso modelo sugere que a topologia da rede pode ter um papel ativo na determinação do desenvolvimento embrionário.

Capítulo 8

Perspectivas Futuras

POR não incluir mecanismos como adesão celular diferenciada, migração celular e apoptose, o nosso modelo fica limitado ao estudo das fases iniciais da embriogênese, quando esses mecanismos não são muito relevantes. Dessa forma, pretendemos levantar essa limitação incluindo esses mecanismos que foram omitidos em primeira aproximação. Provavelmente, para galgarmos a esse nível de abordagem se fará necessária também uma melhor determinação das características dos genes associados ao desenvolvimento como conectividade e genes influenciados por eles.

Algumas modificações devem ser feitas na rede de regulação gênica e na dinâmica da regulação com o intuito de torná-las mais realistas. Entre elas podemos citar: um maior controle do número de genes que se auto-regulam, uma vez que não é um fato que todos os genes possuem auto-regulação, como foi admitido por questão de simplicidade neste trabalho; a atribuição de diferentes pesos às ligações regulatórias, já que nem todos os genes devem ter o mesmo potencial de ativar ou inibir a transcrição; e a consideração de uma dinâmica assíncrona, uma vez que nem todos os genes são afetados pela regulação ao mesmo tempo.

Referências Bibliográficas

- [1] Bruce Alberts *et al.* *Biologia Molecular da Célula*. 3ª ed. Artes Médicas, 1997.
- [2] Maximino Aldana González. Estructura y dinámica de redes genéticas. *Gaceta Biomédicas*, Febrero de 2007.
- [3] http://www.cdcc.usp.br/ciencia/artigos/art_33/EraUmaVez.html. (03/02/2010).
- [4] <http://lexikon.freenet.de/Myokard>. (03/02/2010).
- [5] Peter H. Raven and George B. Johnson. *Biology 6 ed.* McGraw-Hill Higher Education, 2001.
- [6] A. Réka and A. L. Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74:47, 2002.
- [7] F. A. Rodrigues. Caracterização, classificação e análise de redes complexas. *Tese de Doutorado, USP- Instituto de Física de São Carlos*, 2007.
- [8] N. Geard and J. Wiles. A gene network for developing cell lineages. *Artificial Life*, 11:249, 2005.
- [9] A. Réka. Scale-free networks in cell biology. *Journal of Cell Science*, 118:4947, 2005.
- [10] N. M. Luscombe *et al.* Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature*, 431:308, 2004.

- [11] G. T. Dewey. Gene regulatory networks. *Madame Curie Bioscience Database*.
- [12] E. Ravasz and A. L. Barabási. Hierarchical organization in complex networks. *Phys. Rev. E*, 67:026112, 2003.
- [13] S. A. Kauffman. Random boolean network models and the yeast transcriptional network. *PNAS*, 100:14796, 2003.
- [14] E. Domany. Cluster analysis of gene expression data. *J. Stat. Phys.*, 110:1117, 2003.
- [15] S. Bornholdt. Boolean networks models of cellular regulation: prospects and limitations. *J. R. Soc. Interface*, 5:S85, 2008.
- [16] M. I. Davidich and S. Bornholdt. Boolean network model predicts cell cycle sequence of fission yeast. *PLoS ONE*, 3(2):e1672, 2008.
- [17] S. A. Kauffman. *The Origins of Order*. Oxford University Press, 1993.
- [18] I. Shmulevitch and *et al.* Eukaryotic cells are dynamically ordered or critical but not chaotic. *PNAS*, 102:13439, 2005.
- [19] M. Aldana. Boolean dynamics of networks with scale-free topology. *Phys.D.*, 185:45, 2003.
- [20] N. Guelzim *et al.* Topological and causal structure of the yeast transcriptional regulatory network. *Nature Genetics*, 31:60, 2002.
- [21] S. A. Kauffman. Antichaos and adaptation. *Scientific American*, page 64, Agosto-1991.
- [22] N. S. Holter *et al.* Dynamic modeling of gene expression data. *PNAS*, 98:1696, 2001.
- [23] U. Deppe. Cell lineages of the embryo of the nematode *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA*, 75:376, 1978.

- [24] H. S. Silva and M. L. Martins. A cellular automata model for cell differentiation.
Phys. A, 322:555, 2003.