

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

**Caracterização de Caminhos Biológicos na
Dinâmica de Redes Booleanas
Limiarizadas**

Miguel Pereira Ostrowski

MONOGRAFIA FINAL

MAC 499 — TRABALHO DE
FORMATURA SUPERVISIONADO

Supervisor: Prof^a. Dr^a. Ronaldo Fumio Rashimoto
Cossupervisor: Dr^a. Shantanu Gupta

Durante o desenvolvimento deste trabalho o autor recebeu auxílio financeiro da CNPq

São Paulo
2023

*O conteúdo deste trabalho é publicado sob a licença CC BY 4.0
(Creative Commons Attribution 4.0 International License)*

Resumo

Miguel Pereira Ostrowski. **Caracterização de Caminhos Biológicos na Dinâmica de Redes Booleanas Limiarizadas**. Monografia (Bacharelado). Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2023.

As interações entre genes são as responsáveis pela coordenação da execução de funções biológicas, como a divisão celular. Tais interações produzem uma dinâmica que pode ser representada por um diagrama de estados. Nesse diagrama, cada estado possui um único arco para outro estado. Quando um estado faz parte de um ciclo, chama-se esse estado de atrator. Dessa forma, tal diagrama contém muitos caminhos de um estado até um atrator, onde cada caminho pode conter propriedades a depender do grafo de regulação. Este trabalho tem o intuito de estudar características no grafo de regulação que induzem a existência de certas propriedades no diagrama de estado.

Palavras-chave: Redes Booleanas. Genes.

Abstract

Miguel Pereira Ostrowski. **Characterization of Biological Paths in the Dynamics of Thresholded Boolean Networks**. Capstone Project Report (Bachelor). Institute of Mathematics and Statistics, University of São Paulo, São Paulo, 2023.

The interactions between genes are responsible for coordinating the execution of biological functions, such as cell division. These interactions generate a dynamics that can be represented by a state diagram. In this diagram, each state has a single arc to another state. When a state is part of a cycle, it is called an attractor state. Thus, this diagram contains many paths from a state to an attractor, where each path may have properties depending on the regulation graph. This work aims to study characteristics in the regulation graph that induce the existence of certain properties in the state diagram.

Keywords: Boolean Networks. Genes.

Lista de abreviaturas

ABNT	Associação Brasileira de Normas Técnicas
IME	Instituto de Matemática e Estatística
USP	Universidade de São Paulo
I	Iniciais
A1	Ativadores1
A2	Ativadores2
D	Desativadores
F	Finais
C1	Camada1
C2	Camada2

Lista de símbolos

$\phi_i(t)$ Sinal de entrada do gene i no tempo t

Lista de figuras

1.1	a) Grafo da interação gênica; b) Matriz de interação gênica; c) Diagrama de estados gerado.	2
1.2	Rede de interações proposta por Li et al., 2004	4
1.3	Tamanho de cada bacia de atração (primeira coluna) com seu respectivo atrator.	4
1.4	Estados do caminho biológico com sua respectiva fase da divisão celular.	5
1.5	Grafo de regulação da <i>fission yeast</i> . Os arcos amarelos são os arcos de autodegradação, os verdes são de ativação e os vermelhos de degradação. Figura retirada de DAVIDICH e BORNHOLDT, 2013	5
1.6	Caminho biológico da <i>fission yeast</i> . Figura retirada de DAVIDICH e BORNHOLDT, 2013	6
1.7	As trajetórias dos 1.764 estados (nós verdes) fluindo para o ponto fixo (nó azul). Os arcos entre os estados indicam a direção de um estado para outro. A sequência indicada pelos arcos em azul é o caminho biológico. O tamanho de um nó e a espessura de um arco são proporcionais ao logaritmo do número de trajetórias que passa por eles. Figura extraída de Li et al., 2004	7
1.8	Diagrama de estados da <i>fission yeast</i> . Caminho em azul é o caminho biológico da <i>fission yeast</i> . Figura retirada de Li et al., 2004	8
2.1	a) Representação de um gene Tipo 1. Esse gene possui auto degradação e não contém arcos entrando. b) Representação de um gene Tipo 2. Esse gene não possui auto degradação e não contém arcos positivos saindo.	11

2.2	Representação da estrutura que tenta generalizar o comportamento de geração de cascata da <i>budding yeast</i> , de forma que genes Iniciais e Finais serão os únicos nós ligados no primeiro estado do caminho biológico e os genes Finais ligados serão os únicos nós ligados no atrator. Nós com a cor vermelha são nós que possuem auto degradação. Arcos indo de um conjunto de nós A para outro conjunto de nós B, significa que há arcos que saem de nós de A e vão para nós de B. Círculos pretos em um arco que vai de um conjunto de nós para outros, possui o significado que todos os nós daquele conjunto recebe ou manda (a depender da posição que está o círculo) arcos daquele tipo. De forma análoga, um círculo com o meio branco significa que ao menos um nó daquele conjunto recebe ou manda (a depender da posição que está o círculo) arcos do tipo em que o círculo se encontra (arcos de ativação ou de inibição).	13
2.3	Nós Iniciais.	14
2.4	Nós Ativadores1.	14
2.5	Nós Desativadores.	15
2.6	Nós Ativadores2.	15
2.7	Nós da Camada1.	16
2.8	Nós da Camada2.	16
2.9	Nós Finais.	17
3.1	Histograma da razão entre o tamanho da bacia de atração que se encontra o caminho da Conjectura 1 e a quantidade total de estados do diagrama de estados.	21

Lista de tabelas

A.1	Tabela dos estados do caminho biológico e a respectiva fase que cada estado se encontra. Tabela retirada de LEE e HUANG, 2009.	25
-----	--	----

A.2 Tabela A.1 que foi ordenada seguindo o algoritmo A.1. 26

Lista de programas

A.1 Facilita visualização de cascata. 26

Sumário

1	Introdução	1
1.1	Redes Biológicas modeladas por Redes Booleanas	3
1.2	Caminho Biológico	6
1.3	Objetivos	9
2	Metodologia	11
2.1	Definição da estrutura que produz um caminho que respeite a Definição 1	12
2.1.1	Restrições dos nós Iniciais	14
2.1.2	Restrições dos nós Ativadores1	14
2.1.3	Restrições dos nós Desativadores	14
2.1.4	Restrições dos nós Ativadores2	15
2.1.5	Restrições dos nós Camada1	15
2.1.6	Restrições dos nós Camada2	16
2.1.7	Restrições dos nós Finais	16
2.1.8	Restrições adicionais	17
2.2	Ideia de prova da Conjectura 1	17
2.3	Simulações com grafos de regulação	19
3	Resultados	21
3.1	Discussão	21
4	Conclusão	23
4.1	Estudos Futuros	23
Apêndices		
A	Algoritmo que facilita visualização de cascata gerada pelo caminho biológico	25

Anexos

Referências

Capítulo 1

Introdução

Dentro da área de Biologia de Sistemas, muitos modelos matemáticos e computacionais estão sendo desenvolvidos para explicar interações entre seus componentes e existe um número considerável de tentativas para modelar redes biológicas. Um modelo que tem chamado muita atenção é o modelo de redes *Booleanas*, ou *BNs* (*Boolean Networks*), onde utiliza-se grafos direcionados como modelagem a fim de representar uma interação entre genes. Neste grafo, chamado de grafo de interação gênica, cada vértice (ou nó) representa um gene; e um arco liga dois nós se, e somente se, existir uma interação (de ativação ou inibição) do primeiro gene sobre o segundo (ou, analogamente, o segundo gene sofre uma interação do primeiro). Tal grafo, com n genes, pode ser representado por uma matriz de interação gênica $A_{n \times n}$, tal que:

$$a_{i,j} = \begin{cases} 1, & \text{se } x_i \text{ é ativado por } x_j \\ -1, & \text{se } x_i \text{ é inibido por } x_j \\ 0, & \text{se não há interação entre } x_i \text{ e } x_j \end{cases} \quad (1.1)$$

Um exemplo de um grafo de interações e sua matriz correspondente podem ser vistos na Figura 1.1a e 1.1b, respectivamente.

Considerando que cada gene $x_i(t)$ em um determinado tempo t pode assumir valor 1 ou 0 (indicando se está ativado ou desativado, respectivamente), podemos, juntamente com a rede *Booleana*, definir a função $\phi_i(t)$, chamada de sinal de entrada no gene x_i no tempo t , a seguir:

$$\phi_i(t) = \sum_{j=1}^n a_{i,j} \cdot x_j(t) \quad (1.2)$$

Com base na equação podemos calcular no tempo seguinte se um gene é ativado, desativado ou mantém seu estado atual, por meio da seguinte equação:

$$x_i(t+1) = \begin{cases} 1, & \text{se } \phi_i(t) > 0 \\ 0, & \text{se } \phi_i(t) < 0 \\ x_i(t), & \text{se } \phi_i(t) = 0 \end{cases} \quad (1.3)$$

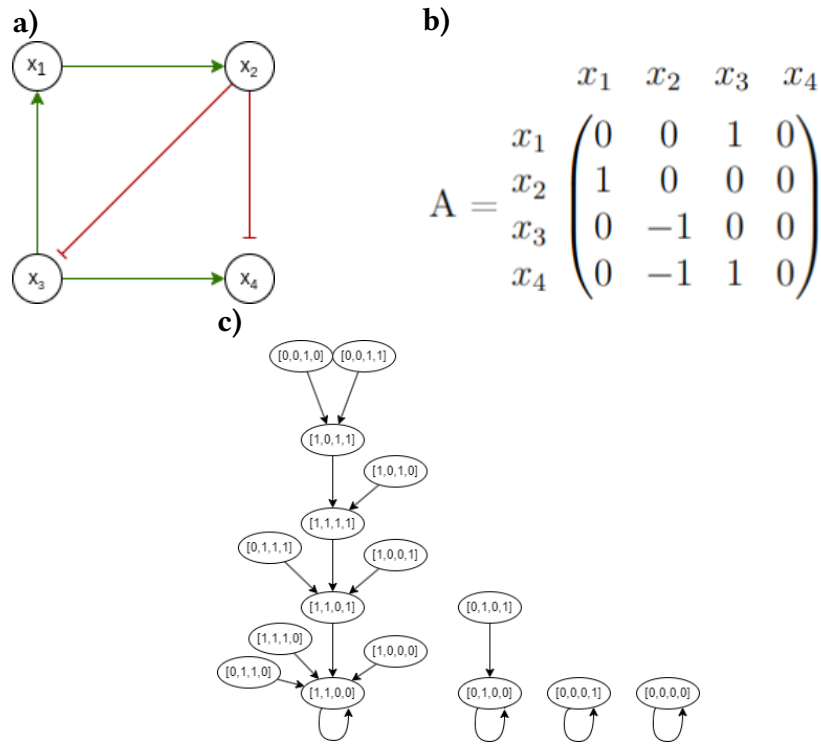


Figura 1.1: a) Grafo da interação gênica; b) Matriz de interação gênica; c) Diagrama de estados gerado.

Como cada gene $x_i(t)$ no tempo t pode assumir somente 2 valores, o vetor $\mathbf{s}(t) = (x_1(t), \dots, x_n(t))$ formado pelos valores binários das n variáveis da *BN* forma uma n -upla pertencente ao conjunto $\{0, 1\}^n$. Definimos, dessa forma, o estado de uma *BN* com n variáveis no tempo t como sendo um vetor binário $\mathbf{s}(t) \in \{0, 1\}^n$. É fácil de ver que uma *BN* com n genes possui 2^n estados possíveis. Dado um estado no tempo t , a atualização de todos os genes no tempo $t + 1$ leva a rede para o próximo estado $\mathbf{s}(t + 1)$. Podemos representar estas transições de estado através de um diagrama de transição de estados. Veja um exemplo na Figura 1.1c.

Uma vez que o número de estados de uma *BN* é finito e que a rede sempre transita de um estado para outro, é fácil ver que depois do sistema iterar por um determinado tempo, certos estados serão revisitados de uma maneira cíclica. Tais estados formam o que chamamos de atratores. Um atrator é um ciclo dirigido (ou um ponto fixo) no diagrama de transição estados da *BN*. Dado um atrator, o conjunto de todos os estados que conduzem a este atrator (inclusive os próprios estados dentro do atrator) é chamado de bacia de atração. A dinâmica de uma *BN* pode ser caracterizada pela quantidade de atratores, pelo tamanho médio dos ciclos atratores, pelo tamanho das bacias de atração, entre outras medidas do diagrama de estados.

1.1 Redes Biológicas modeladas por Redes Booleanas

O ciclo celular da levedura é composto por quatro fases distintas: G_1 (crescimento da célula), S (replicação de DNA), G_2 (preparação para mitose) e M (divisão celular). Neste relatório, nós vamos considerar que o sinal para o início do ciclo celular é o tamanho da célula, indicada pela variável T . As três primeiras fases são conhecidas como intérfase e a última é a mitose. Este ciclo é altamente regulado no sentido de que uma fase não pode ser iniciada sem que a fase anterior tenha sido concluída. Este mecanismo propõe um certo nível de robustez pois as fases devem ser realizadas em sequência para que o ciclo seja bem sucedido. Uma vantagem de se trabalhar com o ciclo celular da levedura é que este processo é relativamente conhecido e bem estudado devido a simplicidade do organismo.

Após uma exaustiva investigação na literatura sobre redes de regulação gênica da *budding yeast*, [Li et al., 2004](#) modelaram o sistema de controle do seu ciclo celular sob o modelo de redes Booleanas e estudaram suas propriedades dinâmicas e estruturais. A rede de regulação gênica proposta por [Li et al., 2004](#) é composta de 11 genes/proteínas representados por x_1, x_2, \dots, x_{11} de forma que $x_i = 1$ e $x_i = 0$ representam o estado ativo e inativo do gene/proteína i , respectivamente. Assim, um estado da rede de [Li et al., 2004](#) é uma configuração do espaço $S = \{0, 1\}^{11}$. Temos então 2^{11} estados, rotulados por inteiros do conjunto $\{0, 1, \dots, 2047\}$.

A rede de interação da *budding yeast* proposta por [Li et al., 2004](#) é mostrada na Figura 1.2. A partir desta figura, é possível construir uma matriz de regulação $A_{11 \times 11}$ da forma descrita na Eq. 1.1: (i) uma flecha de um gene i para um gene j representa que o do gene j é ativado pelo gene i , e neste caso, $a_{ji} = 1$; (ii) um arco com uma barra na ponta de um gene i para um gene j representa que o do gene j é reprimido pelo gene i , e neste caso $a_{ji} = -1$; (iii) laços pontilhados representam autodegradação, e neste caso, $a_{ii} = -1$; (iv) para os casos restantes, $a_{ji} = 0$.

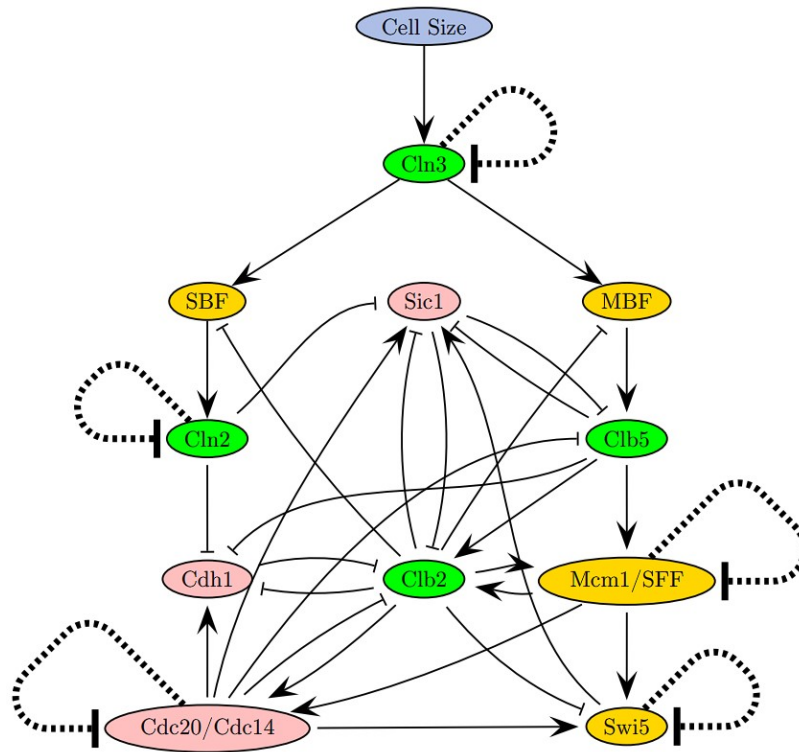


Figura 1.2: Rede de interações proposta por Li et al., 2004.

A dinâmica da rede do Li et al., 2004 tem 7 atratores (todos pontos fixos) e seus tamanhos estão apresentados na tabela abaixo:

Basin Size	Cln3	MBF	SBF	Cln1,2	Cdh1	Swi5	Cdc20	Clb5	Sic1	Clb2	Mcm1
1,764	0	0	0	0	1	0	0	0	1	0	0
151	0	0	1	1	0	0	0	0	0	0	0
109	0	1	0	0	1	0	0	0	1	0	0
9	0	0	0	0	0	0	0	0	1	0	0
7	0	1	0	0	0	0	0	0	1	0	0
7	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	1	0	0	0	0	0	0

Figura 1.3: Tamanho de cada bacia de atração (primeira coluna) com seu respectivo atrator.

A bacia de atração de maior tamanho tem aproximadamente 80% dos estados possíveis e está representada na Figura 1.7. Nesta bacia há uma trajetória com 13 estados (marcada pelos arcos em azul) que representa o ciclo celular da levedura (modela um fenômeno biológico). Esta trajetória de estados é mostrada na tabela abaixo:

State	Cln3	MBF	SBF	Cln2	Cdh1	Swi5	Cdc20	Clb5	Sic1	Clb2	Mcm1	Phase
s_1	1	0	0	0	1	0	0	0	1	0	0	Excited G_1
s_2	0	1	1	0	1	0	0	0	1	0	0	G_1
s_3	0	1	1	1	1	0	0	0	1	0	0	G_1
s_4	0	1	1	1	0	0	0	0	0	0	0	G_1
s_5	0	1	1	1	0	0	0	1	0	0	0	S
s_6	0	1	1	1	0	0	0	1	0	1	1	G_2
s_7	0	0	0	1	0	0	1	1	0	1	1	M
s_8	0	0	0	0	0	1	1	0	0	1	1	M
s_9	0	0	0	0	0	1	1	0	1	1	1	M
s_{10}	0	0	0	0	0	1	1	0	1	0	1	M
s_{11}	0	0	0	0	1	1	1	0	1	0	0	M
s_{12}	0	0	0	0	1	1	0	0	1	0	0	G_1
s_{13}	0	0	0	0	1	0	0	0	1	0	0	Stationary G_1

Figura 1.4: Estados do caminho biológico com sua respectiva fase da divisão celular.

Observe que cada estado na trajetória corresponde a uma fase da divisão celular da levedura da *budding yeast*.

Adiante, outra rede de regulação gênica conhecida modela o ciclo celular da *fission yeast*. Essa rede de regulação possui 12 genes/proteínas representados por y_1, y_2, \dots, y_{12} de modo que $y_i = 1, y_i = 0$ representam respectivamente o estado ativo e inativo do gene/proteína i . Dessa forma, um estado da rede da *fission yeast* é uma configuração do espaço $T = 0,1^{12}$. Portanto, essa rede gera um diagrama de estados com 2^{12} estados.

A rede de interação gênica da *fission yeast* está exposta na Figura 1.5. Com essa figura é possível formar uma matriz de regulação $B_{12 \times 12}$ como definido na Eq. 1.1, de modo que (i) um arco verde de um gene i até um gene j representa que o gene j é ativado pelo gene i ; (ii) um arco vermelho de um gene i até um j representa que um gene j é reprimido pelo gene i ; (iii) laços amarelos representam que um gene i se reprime; (iv) para os casos restantes, não há interação entre os genes.

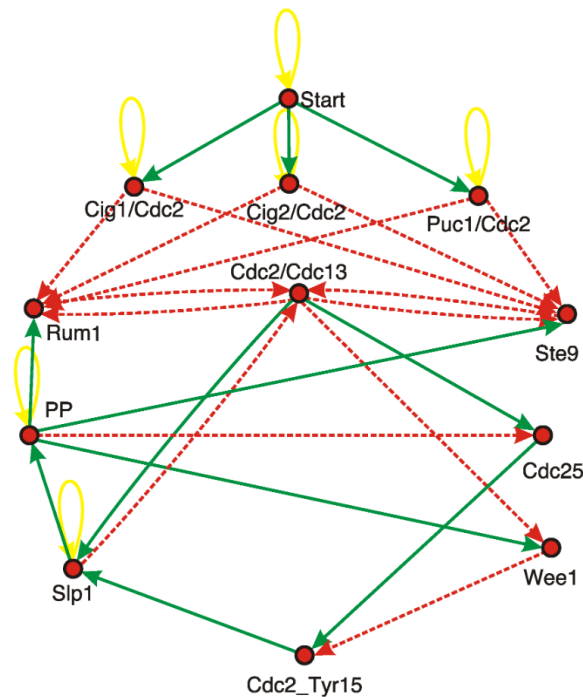


Figura 1.5: Grafo de regulação da *fission yeast*. Os arcos amarelos são os arcos de autodegradação, os verdes são de ativação e os vermelhos de degradação. Figura retirada de DAVIDICH e BORNHOLDT, 2013.

A dinâmica gerada pela *fission yeast* possui um caminho com 10 estados, onde cada estado possui uma descrição de cada fase do ciclo celular, como pode ser visto na Figura 1.6.

Time Step	Start	SK	Cdc2/Cdc13	Ste9	Rum1	Slp1	Cdc2/Cdc13*	Wee1 Mik1	Cdc25	PP	Phase	comments
1	1	0	0	1	1	0	0	1	0	0	START	Cdc2/Cdc13 dimers are inhibited, antagonists are active.
2	0	1	0	1	1	0	0	1	0	0	G1	SK are becoming active
3	0	0	0	0	0	0	0	1	0	0	G1/S	When Cdc2/Cdc13 and SK dimers switch off Rum1 and Ste9/APC, the cell passes 'Start' and DNA replication takes place, Cdc2/Cdc13 starts to accumulate
4	0	0	1	0	0	0	0	1	0	0	G2	Activity of Cdc2/Cdc13 achieves moderate level, which is enough for entering G2 phase but not mitosis, since Wee1/Mik1 inhibits the activity of residue Tyr-15 of Cdc2 (Cdc2/Cdc13* is not active)
5	0	0	1	0	0	0	0	0	1	0	G2	Moderate activity Cdc2/Cdc13 activates Cdc25
6	0	0	1	0	0	0	1	0	1	0	G2/M	Cdc25 reverses phosphorylation, removing the inhibiting phosphate group and activating Cdc2/Cdc13*
7	0	0	1	0	0	1	1	0	1	0	G2/M	Cdc2/Cdc13* reaches high activity level sufficient to activate Slp1/APC mitosis
8	0	0	0	0	0	1	0	0	1	1	M	Slp1 degrades Cdc13, that is inhibits complex Cdc2/Cdc13 and Cdc2/Cdc13*.
9	0	0	0	1	1	0	0	1	0	1	M	Antagonists of Cdc2/Cdc13 are reset.
10	0	0	0	1	1	0	0	1	0	0	G1	Cell reaches G1 stationary state (PP is inactive)

doi:10.1371/journal.pone.0001672.t002

Figura 1.6: Caminho biológico da *fission yeast*. Figura retirada de *DAVIDICH e BORNHOLDT, 2013*.

1.2 Caminho Biológico

Com as fórmulas descritas anteriormente, podemos aplicá-las em um grafo de regulação conhecido, como o grafo de regulação da *budding yeast*, de modo que é calculado a transição de estado para cada combinação de genes ligados e desligados. Dessa forma obtêm-se o diagrama de estados da *budding yeast*, onde a maior bacia de atração desse diagrama está representado Figura 1.7.

O caminho biológico da *budding yeast*, descrito na tabela da Figura 1.4, pode ser observado no caminho em azul na Figura 1.7.

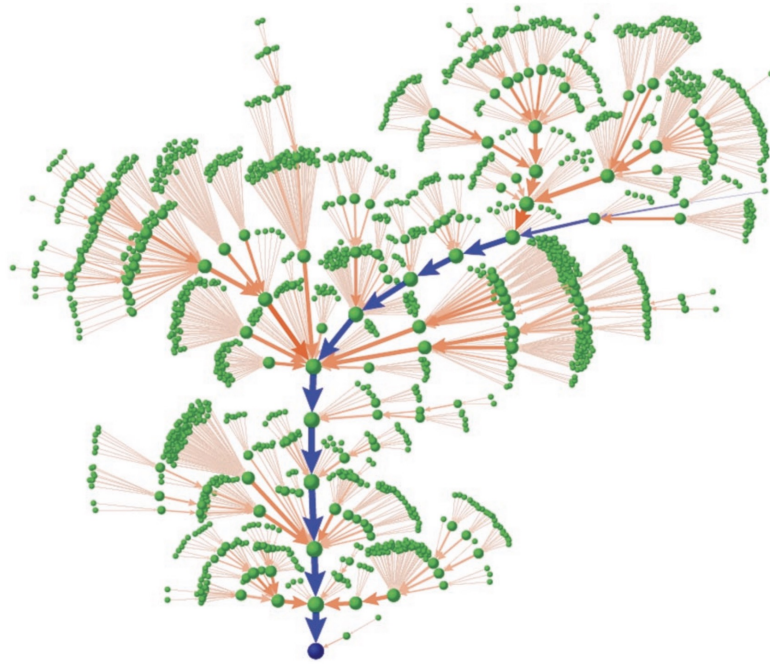


Figura 1.7: As trajetórias dos 1.764 estados (nós verdes) fluindo para o ponto fixo (nó azul). Os arcos entre os estados indicam a direção de um estado para outro. A sequência indicada pelos arcos em azul é o caminho biológico. O tamanho de um nó e a espessura de um arco são proporcionais ao logaritmo do número de trajetórias que passa por eles. Figura extraída de *Li et al., 2004*.

Bem como o caminho biológico da *fission yeast*, descrito na tabela da Figura 1.6, pode ser verificado no caminho em azul da Figura 1.8.

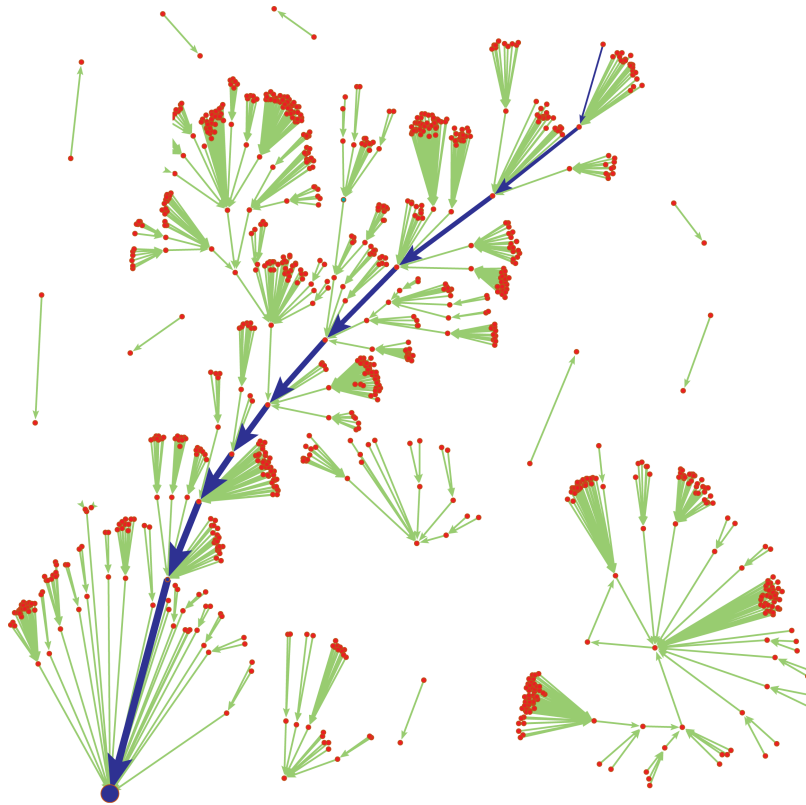


Figura 1.8: Diagrama de estados da fission yeast. Caminho em azul é o caminho biológico da fission yeast. Figura retirada de Li et al., 2004.

Como é possível verificar na Figura 1.7, há um caminho em azul que se destaca. Esse é o caminho biológico, que modela o ciclo celular da levedura da *budding yeast*. Cada estado desse caminho possui uma descrição na Figura 1.4.

Agora, pode-se verificar que os caminhos biológicos da *budding yeast* e da *fission yeast* possuem a seguinte definição:

Definição 1. O primeiro e o último estado do caminho são respectivamente uma folha e um ponto fixo. Além disso, os genes ligados no último estado também estão ligados no primeiro estado, ou seja, o conjunto de genes ligados no último estado está contido no conjunto de genes ligados do primeiro estado.

A Definição 1 pode ser verificada para a *budding yeast*, com a tabela da Figura 1.6, onde os genes Cln3, Cdh1 e o Sic1 estão ligados no primeiro estado e apenas o Cdh1 e o Sic1 estão ligados no último estado. Além disso, é possível verificar na Figura 1.7 que o primeiro e o último estado do caminho biológico da *budding yeast* são respectivamente uma folha e um ponto fixo. Portanto o caminho biológico da *budding yeast* está de acordo com a Definição 1.

De forma análoga, também pode-se verificar a Definição 1 para a *fission yeast*, onde pode-se utilizar a tabela da Figura 1.6 para verificar que o gene Start, o gene Ste9, o gene Rum1 e o gene Wee1 estão ligados no primeiro estado, e somente os genes Ste9, Rum1 e o Wee1 estão ligados no último estado. Ademais, é possível observar na Figura 1.8 que o

primeiro e o último estado do caminho biológico da *fission yeast* são respectivamente uma folha e um ponto fixo.

Entretanto, a existência de um caminho que respeite a Definição 1 não pode ser esperada para qualquer grafo de regulação, uma vez que a existência desse caminho depende da estrutura do grafo de regulação.

1.3 Objetivos

Este trabalho tem o objetivo de estudar a dinâmica gerada pelos grafos de regulação da *budding yeast* e da *fission yeast*, com o intuito de definir um conjunto de grafos de regulação que contenham um caminho que respeite a Definição 1.

Capítulo 2

Metodologia

Com o intuito de estudar como os grafos de regulação da *budding yeast* e da *fission yeast* geram um caminho que respeite a Definição 1, são feitas observações nos grafos de regulação das leveduras, com a intenção de encontrar semelhanças em seus grafos de regulação que produzam um caminho que respeite a Definição 1.

Como resultado dessas observações, pode-se afirmar que é possível segmentar os genes ligados no primeiro estado do caminho biológico dessas leveduras em genes do Tipo 1 e genes do Tipo 2. Os genes do Tipo 1 são definidos como genes que possuem auto-degradação e não contém arcos vindos de outros genes, enquanto os genes do Tipo 2 são genes que não possuem auto-degradação, além de não conter arcos de ativação para outros genes. Podemos ver a representação de um gene Tipo 1 e de um gene Tipo 2 na Figura 2.1, a seguir:

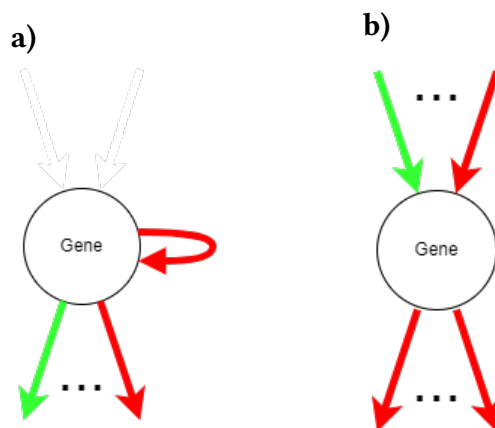


Figura 2.1: a) Representação de um gene Tipo 1. Esse gene possui auto degradação e não contém arcos entrando. b) Representação de um gene Tipo 2. Esse gene não possui auto degradação e não contém arcos positivos saindo.

Também é observado que os únicos genes ligados no último estado do caminho biológico das duas leveduras são genes do Tipo 2. Essas características podem ser verificadas para a *budding yeast* com o grafo de regulação da Figura 1.2, de modo que o gene o Cln3 é um gene do Tipo 1 e os genes Cdh1 e o Sic1 são os genes do Tipo 2. Com isso, é possível

verificar na descrição do caminho biológico da Figura 1.4 que a *budding yeast* possui o primeiro estado do caminho biológico contendo somente genes do Tipo 1 e genes do Tipo 2 ligados, enquanto o último estado possui somente genes do Tipo 2 ligados.

De forma análoga, também pode-se verificar esse comportamento no grafo de regulação da *fission yeast*, na qual é possível utilizar seu grafo de regulação, apresentado na Figura 1.5, para constatar que o gene Start é um gene do Tipo 1 e os genes Ste9, Rum1 e o Wee1 são genes do Tipo 2. Assim, é possível observar no caminho biológico da *fission yeast*, representado na Figura 1.6, que o primeiro estado contém somente genes do Tipo 1 e genes do Tipo 2 ligados, enquanto o último estado possui somente genes do Tipo 2 ligados.

2.1 Definição da estrutura que produz um caminho que respeite a Definição 1

Com o intuito de criar uma estrutura de grafo de regulação que possua um caminho que respeite a Definição 1, é proposto uma estrutura inspirada nos grafos de regulação da *budding yeast* e da *fission yeast* e também em teoremas de [TRAN *et al.*, 2013](#), que é representada na Figura 2.2 a seguir:

2.1 | DEFINIÇÃO DA ESTRUTURA QUE PRODUZ UM CAMINHO QUE RESPEITE A DEFINIÇÃO 1

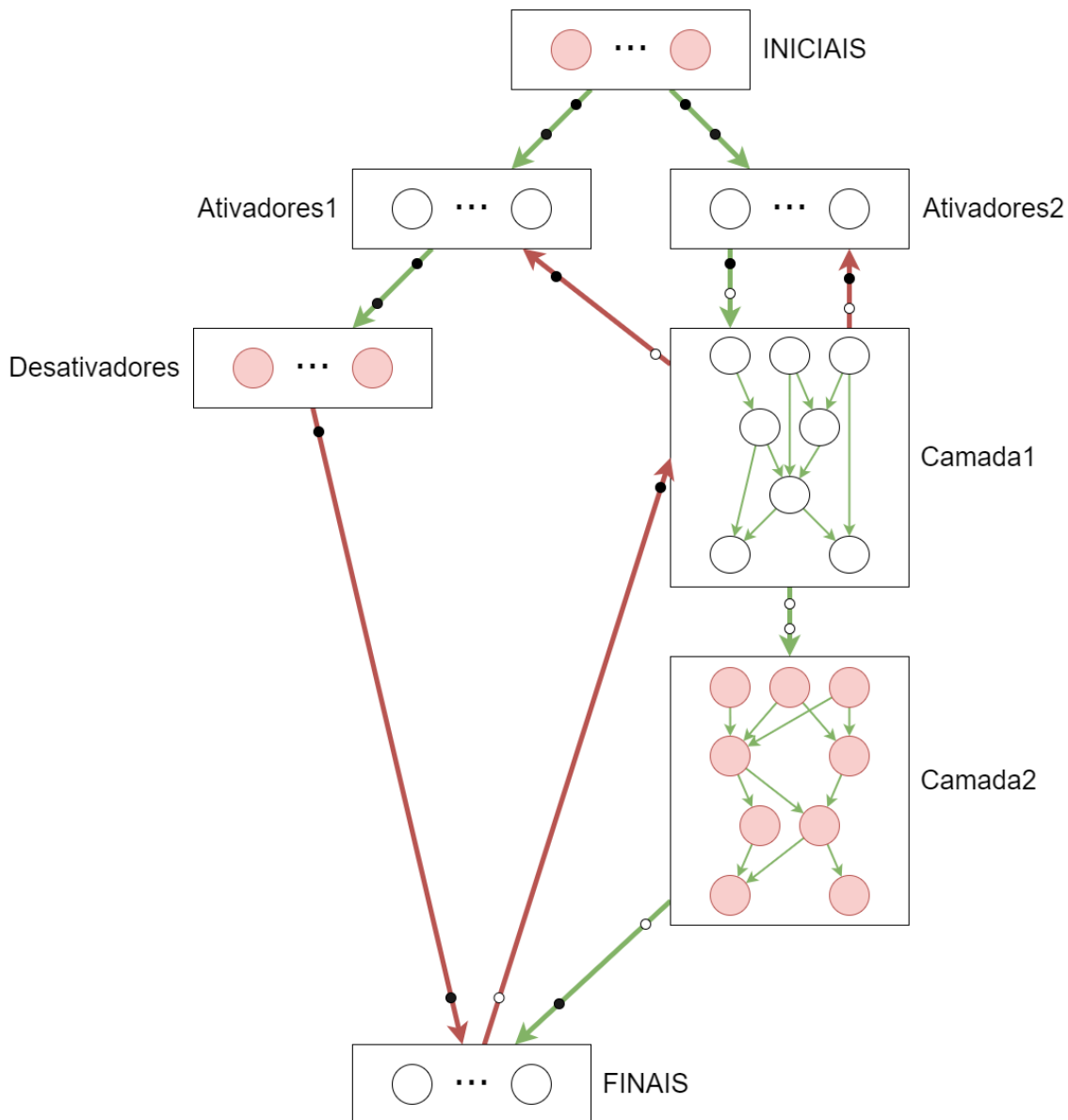


Figura 2.2: Representação da estrutura que tenta generalizar o comportamento de geração de cascata da budding yeast, de forma que genes Iniciais e Finais serão os únicos nós ligados no primeiro estado do caminho biológico e os genes Finais ligados serão os únicos nós ligados no atrator. Nós com a cor vermelha são nós que possuem auto degradação. Arcos indo de um conjunto de nós A para outro conjunto de nós B, significa que há arcos que saem de nós de A e vão para nós de B. Círculos pretos em um arco que vai de um conjunto de nós para outros, possui o significado que todos os nós daquele conjunto recebe ou manda (a depender da posição que está o círculo) arcos daquele tipo. De forma análoga, um círculo com o meio branco significa que ao menos um nó daquele conjunto recebe ou manda (a depender da posição que está o círculo) arcos do tipo em que o círculo se encontra (arcos de ativação ou de inibição).

Todas as restrições de cada parte da estrutura, representada na Figura 2.2, são discutidas nas próximas seções.

2.1.1 Restrições dos nós Iniciais

Os nós contidos em I (Iniciais), que estão representados na Figura 2.3), são genes do Tipo 1, dessa forma os genes contidos em I possuem as seguintes restrições:

1. $\forall u \in I, a_{uu} = -1$.
2. $\forall u \in I, \exists v \in A1$ tal que $a_{vu} = 1$
3. $\forall u \in I, \exists v \in A2$ tal que $a_{vu} = -1$
4. $\forall u, v \in I$, se $u \neq v$ então $a_{vu} = 0$.

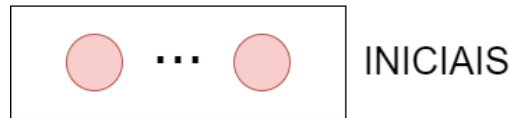


Figura 2.3: Nós Iniciais.

2.1.2 Restrições dos nós Ativadores1

Os nós contidos em $A1$ (Ativadores1), que estão representados na Figura 2.4) contém as seguintes restrições:

1. $\forall u \in A1, \exists v \in D$ tal que $a_{vu} = 1$
2. $\forall u \in A1, \exists v \in I$ tal que $a_{uv} = 1$
3. $\forall u \in A1, \exists v \in C1$ tal que $a_{uv} = -1$
4. $\forall u, v \in A1, a_{vu} = 0$.

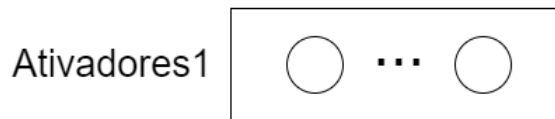


Figura 2.4: Nós Ativadores1.

2.1.3 Restrições dos nós Desativadores

Os nós contidos em D (Desativadores), que estão representados na Figura 2.5) contém as seguintes restrições:

1. $\forall u \in D, a_{uu} = -1$.
2. $\forall u \in D, \exists v \in F$ tal que $a_{vu} = -1$
3. $\forall u \in D, \exists v \in A1$ tal que $a_{uv} = 1$
4. $\forall u, v \in D$ se $u \neq v$ então, $a_{vu} = 0$.

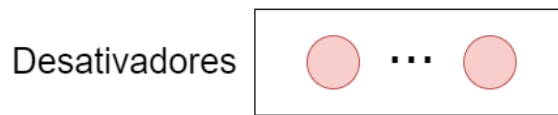


Figura 2.5: Nós Desativadores.

2.1.4 Restrições dos nós Ativadores2

Os nós contidos em $A2$ (Ativadores2), que estão representados na Figura 2.6) contém as seguintes restrições:

1. $\forall u \in A2, \exists v \in C1$ tal que $a_{vu} = 1$
2. $\forall u \in A2, \exists v \in I$ tal que $a_{uv} = 1$
3. $\forall u \in A2, \exists v \in C1$ tal que $a_{uv} = -1$
4. $\forall u, v \in A2, a_{vu} = 0$.

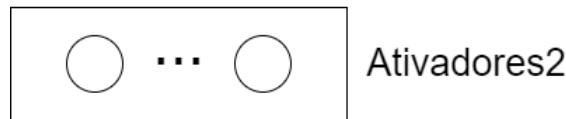


Figura 2.6: Nós Ativadores2.

2.1.5 Restrições dos nós Camada1

Os nós contidos em $C1$ (Camada1) fazem parte de uma Hierarquia, onde define-se uma Hierarquia a seguir:

Definição 2. Uma Hierarquia H é definida como um conjunto de nós, onde $\forall h \in H$ há um inteiro Alt_h relacionada a cada nó, de forma que $Alt_h \geq 0$. Além disso, se $a_{h_1, h_2} = 1$, onde $h_1, h_2 \in H$, então $Alt_{h_1} > Alt_{h_2}$. Por último, para toda hierarquia H , $\exists u \in H$ tal que $Alt_u = 0$.

Os nós contidos em $C1$, que estão representados na Figura 2.7) contém as seguintes restrições:

1. $\forall u \in C1, a_{uu} = 0$.
2. Se $u \in C1$ então $u \in H_1$ onde H_1 é uma Hierarquia, definida em 2.
3. Se Alt_u é máximo em H_1 , $u \in H_1$, então há exatamente um arco positivo que entra nele vindo de um nó contido em $A2$.
4. Se $Alt_u = 0$, $u \in H_1$, então $\exists v \in C2$ tal que $a_{vu} = 1$.
5. $\forall u \in C1, \exists v \in F$ tal que $a_{uv} = -1$

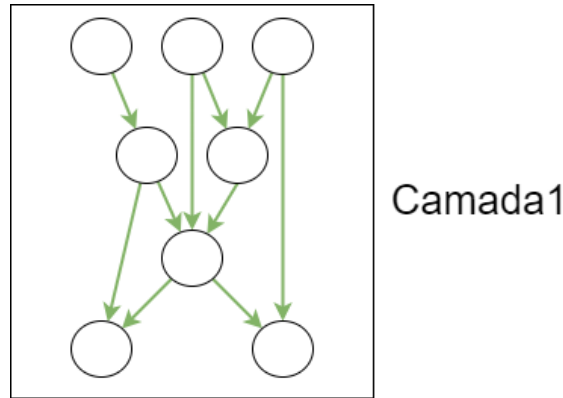


Figura 2.7: Nós da Camada1.

2.1.6 Restrições dos nós Camada2

Os nós contidos em $C2$ (Camada2), que estão representados na Figura 2.8) contém as seguintes restrições:

1. $\forall u \in C2, a_{uu} = -1$.
2. Se $u \in C2$ então $u \in H_2$ onde H_2 é uma Hierarquia, definida em 2.
3. Se Alt_u é máximo, $u \in H_2$, então $\exists v \in C1$ tal que $a_{uv} = 1$.
4. Se $Alt_u = 0, u \in H_2$, então $\exists v \in F$ tal que $a_{vu} = 1$.

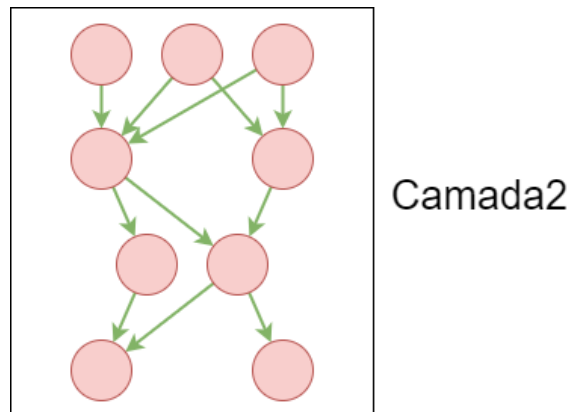


Figura 2.8: Nós da Camada2.

2.1.7 Restrições dos nós Finais

Os nós contidos em F (Finais), representados na Figura 2.9), são genes do Tipo 2, dessa forma os genes contidos em F possuem as seguintes restrições:

1. $\forall u \in F, a_{uu} = -1$.
2. $\exists u \in F, \exists v \in C1$ tal que $a_{vu} = 1$
3. $\forall u \in F, \exists v \in C2$ tal que $a_{uv} = 1$

4. $\forall u \in F, \exists v \in D$ tal que $a_{uv} = 1$
5. $\forall u, v, u \neq v \in F, a_{vu} = 0$.

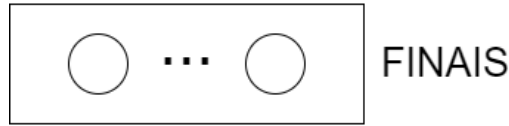


Figura 2.9: Nós Finais.

2.1.8 Restrições adicionais

Além das restrições citadas nas seções anteriores, também há mais duas restrições que a estrutura requiere:

1. \forall nó $a \in I \cup A1 \cup D$, há um caminho, que passe somente por arcos positivos, que parta de um nó de I e termine em um dos nós de D e que contenha a .
2. \forall nó $b \in I \cup A2 \cup C1 \cup C2 \cup F$, há um caminho, que passe somente por arcos positivos, que parta de um nó de I e termine em um nó de F e que contenha b .

Agora, com todas partes da estrutura definida, é proposto a seguinte conjectura:

Conjectura 1. *Se um grafo de regulação obedece as restrições impostas pela estrutura definida na Seção 2.1, então há um caminho, em seu diagrama de estados, que possui a Definição 1.*

2.2 Ideia de prova da Conjectura 1

As duas definições a seguir são necessárias para a ideia de prova:

Definição 3. $\max(\text{Dist}(H))$ é a maior distância que um nó $h_1 \in H$ consegue chegar em um outro nó h_2 , tal que $\text{Alt}_{h_2} = 0$ e Alt_{h_1} é máximo.

Definição 4. $\min(\text{Dist}(H))$ é a menor distância que um nó $h_1 \in H$ consegue chegar em um outro nó h_2 , tal que $\text{Alt}_{h_2} = 0$ e Alt_{h_1} é máximo.

Agora, é exposto a ideia de prova da Conjectura 1:

Primeiro, assume-se que no tempo $t=0$ todos os genes em Iniciais e em Finais estão ligados e que todos os outros genes estão desligados.

1. No tempo $t = 0$.

Ativadores possuem $\phi_q(t) > 0, \forall q \in A1 \cup A2$, uma vez que apenas nos em I e F estão ligados e $\phi_i(t) = -1 \forall i \in I$, logo esses serão desligados.

2. No tempo $t = 1$.

$\forall i \in I$ estão desligados. $\forall a \in A1 \cup A2$ estão ligados, portanto $\phi_a(t) > 0 \forall d \in D$ e $\phi_{c1}(t) \leq 0 \forall c1 \in C1$, já que os genes de F estão ligados.

3. No tempo $t = 2$.

Todos nós em D estão ligados, logo $\phi_f(t) < 0 \forall f \in F$.

4. No tempo $t = 3$.

Nós em F estão todos desligados, dessa maneira $\phi_{c1}(t) \geq 0 \forall c1 \in C1$.

5. No tempo $t = 4$.

Os nós $c1$ que Alt_{c1} é máximo para $C1$ são ligados, dessa maneira esses nós começam a interagir positivamente com outros nós de $C1$. Além disso, nós de $A1$ e $A2$ podem começar a desligar uma vez que os nós de $C1$ estão começando a serem ligados.

6. No tempo $t = 4 + \min(Dist(C1))$.

$\exists c1 \in C1$ tal que $Alt(c1) = 0$ e que $c1$ está ligado, portanto esses nós começam a interagir positivamente com os nós de $C2$, portanto $\exists c2 \in C2$ tal que $\phi_{c2}(t) > 0$.

7. Até $t = 4 + \max(Dist(C1))$.

$\forall c1 \in C1$, $c1$ está ligado, logo até o passo seguinte é garantido que todos os nós em $A1 \cup A2$ estarão desligados, uma vez que os nós de $C1$ estão interagindo negativamente com esses nós. Além disso, se $Alt(c2)$ é máxima em $C2$, então $\phi_{c2}(t) > 0$, já que todos os nós de $C1$ estão ligados e interagindo positivamente com os nós de $C2$ que a $Alt(c2)$ é máxima.

8. Até $t = 4 + \max(Dist(C1)) + 1$.

q esta desligado, \forall nó $q \in A1 \cup A2$, logo $\phi_d(t) < 0 \forall d \in D$, uma vez que nenhum nó de $A1$ está interagindo positivamente com os nós de D e por todos os nós de D terem autodegradação.

9. Até $t = 4 + \max(Dist(C1)) + 2$.

d está desligado, \forall nó $d \in D$.

10. Até $t = 4 + \min(Dist(C1)) + \min(Dist(C2))$.

$\exists c2 \in C2$ tal que $Alt(c2) = 0$ e que $c2$ está ligado, logo $\exists f \in F$ tal que $\phi_f(t) > 0$.

11. Até $t = 4 + \min(Dist(C1)) + \min(Dist(C2)) + 1$.

$\exists f \in F$ tal que f está ligado, logo pode haver $\phi_{c1}(t) < 0$ tal que $c1 \in F$.

12. Até $t = 4 + \max(Dist(C1)) + \max(Dist(C2))$.

$c2$ está ligado, $\forall c2 \in C2$, logo $\phi_f(t) > 0 \forall f \in F$.

13. Até $t = 4 + \max(Dist(C1)) + \max(Dist(C2)) + 1$.

f está ligado, $\forall f \in F$, portanto $\exists c1 \in C1$ tal que $\phi_{c1}(t) < 0$, logo é garantido que os nós da $C1$ começam a ser desligados, uma vez que os nós de F interagem negativamente com os nós de $C1$.

14. Até $t = 4 + 2 * \max(Dist(C1)) + \max(Dist(C2)) + 1$.

$c1$ está desligado, $\forall c1 \in C1$, logo $\exists c2 \in C2$ tal que $\phi_{c2}(t) < 0$.

15. Até $t = 4 + 2 * \max(\text{Dist}(C1)) + 2 * \max(\text{Dist}(C2)) + 1$.

$c2$ está desligado, $\forall c2 \in C2$.

Ao final da prova, todos os nós em F estão ligados, enquanto todos os outros nós estão desligados.

Entretanto, como a prova não está completa, são feitas simulações com o intuito de validar a Conjectura 1, onde os detalhes são descritos na próxima seção.

2.3 Simulações com grafos de regulação

Com o objetivo de validar a Conjectura 1, são feitas 250 simulações com grafos de regulação aleatórios que respeitam as restrições da estrutura do grafo de regulação proposto na Seção 2.1. De modo que I tem o máximo de 2 nós, $A1$ tem um máximo de 2, $A2$ tem um máximo de 2 nós, D possui no máximo de 2 nós, $C1$ contém no máximo 7 nós, $C2$ contém no máximo 6 nós e F possui no máximo 3 nós, ou seja, o maior grafo de regulação gerado pode ter até 24 genes. Assim é testado para cada grafo de regulação gerado se a Conjectura 1 é verdadeira.

Contudo, não foi escolhido uma quantidade maior de simulações, uma vez que o tempo de verificação para cada grafo de regulação cresce exponencialmente com a quantidade de nós do grafo.

Capítulo 3

Resultados

Os resultados indicam que grafos de regulação que seguem as restrições da estrutura proposta na Seção 2.1 possuem um caminho que respeita a Definição 1, como a prova sugere. Entretanto, a partir da simulação com 250 grafos de regulação que respeitam as restrições impostas na Seção 2.1, observa-se que todos os grafos gerados possuem um caminho que respeite a Definição 1, o que demonstra que a Conjectura 1 é válida.

Além disso, também é observado nos testes que o caminho citado na Conjectura 1 sempre esteve em uma bacia de atração grande, como pode ser visto na Figura 3.1. Porém esse resultado não era esperado e foi obtido com base nos testes feitos.

Histograma do tamanho relativo da bacia de atração que contém o caminho

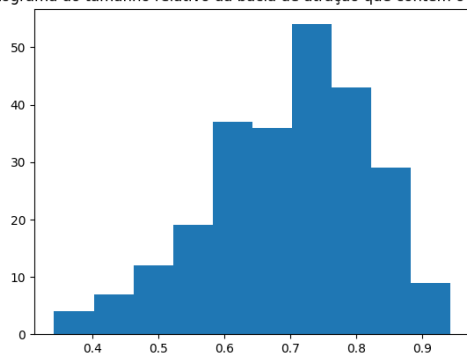


Figura 3.1: Histograma da razão entre o tamanho da bacia de atração que se encontra o caminho da Conjectura 1 e a quantidade total de estados do diagrama de estados.

3.1 Discussão

Os resultado podem ser ser explicados parcialmente por meio dos genes Iniciais e Finais, uma vez que esses genes são, em parte, responsáveis por garantir que exista um caminho que respeite a Definição 1. Isso se deve ao fato de que todos os estados que têm um gene Inicial ligado serão uma folha, uma vez que por possuir auto-degradação e por não possuir arcos vindos de outros genes, então esses estados não podem ficar ligado por dois estados consecutivos. Portanto se um estado possui um gene Inicial ligado, necessariamente esse

estado é uma folha. De forma análogo, não só os genes Finais como também a estrutura definida na Seção 2.1 contribuem para que o último estado desse caminho seja um atrator, uma vez os genes Finais não podem ligar outros genes e a estrutura garante que somente os genes Finais estejam ligados no do caminho.

Capítulo 4

Conclusão

As observações feitas na *budding yeast* e na *fission yeast* se mostraram relevantes para definir um estrutura de grafo de regulação que respeitasse a Definição 1. Isso se deve ao fato que genes que podem ser caracterizados como genes do Tipo 1 e genes do Tipo 2 são parcialmente responsáveis por garantir que esse caminho ocorra.

A partir disso, a estrutura, sugerida na Seção 2.1, se mostrou eficaz em possuir um caminho que contém a Definição 1, de forma a conseguir garantir a existência desse caminho no diagrama de estados para 250 grafos de regulação que seguem as restrições da estrutura definida na Seção 2.1, demonstrando que a Conjectura 1 é possivelmente válida.

Além disso, um resultado que não era esperado dessa estrutura, foi que o caminho que respeita a Definição 1 sempre esteve em uma bacia de atração com tamanho considerável, como pode ser visto no histograma da Figura 3.1.

Entretanto, para que haja um resultado mais robusto é necessário terminar a prova da Seção 2.2 para que possa-se afirmar com certeza que se um grafo segue todas as restrições impostas na estrutura definida na Seção 2.1 então esse grafo possui a Definição 1, mesmo que as 250 simulações não tenham encontrado um contra-exemplo.

4.1 Estudos Futuros

Na literatura é estudada propriedades das interações gênicas que afetam a dinâmica da rede. Como em BORNHOLDT, 2008 afirma, arcos de auto degradação são importantes, em modelos estocásticos, para evitar oscilações na dinâmica. Além disso, em THIEFFRY, 2007 é verificado que, em modelos *Booleanos*, a quantidade de ciclos negativos possui uma correlação com a quantidade de pontos fixos. Dado isso, pode-se utilizar estudos semelhantes a fim de generalizar a estrutura definida na Seção 2.1 com o intuito de aumentar o conjunto de grafos de regulação que possuem um caminho que respeite a Definição 1.

Adiante, segundo LI *et al.*, 2004, a levedura estudada possui a propriedade de que o caminho biológico está na maior bacia de atração do diagrama de estados, e por isso

pequenas perturbações nos estados do caminho biológico podem ter uma alta chance de voltar ao caminho biológico, assim continuando seu processo biológico normalmente. Logo, pode-se estudar quais são as propriedades de um grafo de regulação pode gerar um diagrama de estados com uma bacia de atração com a maioria dos estados, já que neste trabalho, não é explicado como que a estrutura definida na Seção 2.1 obteve esse comportamento como resultado.

Apêndice A

Algoritmo que facilita visualização de cascata gerada pelo caminho biológico

Uma característica interessante produzida pelo caminho biológico é a sequência de estados que são ligados e depois desligados, dando uma ideia de cascata, como pode ser visto na Tabela A.2. Entretanto, dependendo de como as colunas estão posicionadas, não é possível verificar tão facilmente essa cascata, como pode ser visto na Tabela A.1. Com esse objetivo foi feito um algoritmo que ordena as colunas de forma a facilitar essa visualização. O algoritmo utilizado simplifica o problema de forma que cada coluna é representada por uma tupla $\langle a, b, c \rangle$, onde c é o número da coluna; a é o número da linha que tem o primeiro valor não nulo na coluna c ; e b é o maior número da linha que podemos alcançar, partindo de a , passando somente por 1's na coluna c . De forma que, na ordenação, a ordem de maior prioridade para a menor prioridade é a, b, c .

Time	Cln3	MBF	SBF	Cln1,2	Cdh1	Swi5	Cdc20&14	Clb5,6	Sic1Cl	Clb1,2	Mcm1/SFF	Phase
1	1	0	0	0	1	0	0	0	1	0	0	START
2	0	1	1	0	1	0	0	0	1	0	0	G1
3	0	1	1	1	1	0	0	0	1	0	0	G1
4	0	1	1	1	0	0	0	0	0	0	0	G1
5	0	1	1	1	0	0	0	1	0	0	0	S
6	0	1	1	1	0	0	0	1	0	1	1	G2
7	0	0	0	1	0	0	1	1	0	1	1	M
8	0	0	0	0	0	1	1	0	0	1	1	M
9	0	0	0	0	0	1	1	0	1	1	1	M
10	0	0	0	0	0	1	1	0	1	0	1	M
11	0	0	0	0	1	1	1	0	1	0	0	M
12	0	0	0	0	1	1	0	0	1	0	0	G1
13	0	0	0	0	1	0	0	0	1	0	0	Stationary G1

Tabela A.1: Tabela dos estados do caminho biológico e a respectiva fase que cada estado se encontra. Tabela retirada de LEE e HUANG, 2009.

Contudo, esse algoritmo possui uma restrição uma vez que a ordenação é garantida somente para a primeira sequência de 1's de cada coluna, de modo que quando há empate entre duas tuplas, as demais possíveis sequências de 1's dessas colunas, não terão ordenação

Time	Cln3	Sic1Cl	Cdh1	MBF	SBF	Cln1,2	Clb5,6	Mcm1/SFF	Clb1,2	Cdc20&Cdc14	Swi5	Phase
1	1	1	1	0	0	0	0	0	0	0	0	START
2	0	1	1	1	1	0	0	0	0	0	0	G1
3	0	1	1	1	1	1	0	0	0	0	0	G1
4	0	0	0	1	1	1	0	0	0	0	0	G1
5	0	0	0	1	1	1	1	0	0	0	0	S
6	0	0	0	1	1	1	1	1	1	0	0	G2
7	0	0	0	0	0	1	1	1	1	1	0	M
8	0	0	0	0	0	0	0	1	1	1	1	M
9	0	1	0	0	0	0	0	1	1	1	1	M
10	0	1	0	0	0	0	0	1	0	1	1	M
11	0	1	1	0	0	0	0	0	0	1	1	M
12	0	1	1	0	0	0	0	0	0	0	1	G1
13	0	1	1	0	0	0	0	0	0	0	0	Stationary G1

Tabela A.2: Tabela A.1 que foi ordenada seguindo o algoritmo A.1.

garantida. De todo modo, é suficiente para o problema, já que devido a estrutura de interação gênica, cada coluna deve possuir no máximo duas sequências de 1's, logo é bom o suficiente para este problema.

Programa A.1 Facilita visualização de cascata.

```

1  FUNCAO OrdenaMatriz(Matriz, NumLinhas, NumColunas) ▷ Ordena as colunas da
    Matriz de forma a deixar a cascata mais aparente
2
3  MinHeap ← MinHeap()
4  para j = 1 até NumColunas
5      iniciou ← false
6      a ← NumLinhas
7      b ← NumLinhas
8      para i = 1 até NumLinhas
9          se iniciou E Matriz[i, j] = 0
10             b ← i-1
11             interrompe
12         senão se não iniciou E Matriz[i, j] = 1
13             a ← i
14             iniciou ← true
15     fim
16     c ← j
17     MinHeap.Add({a, b, c})
18 fim
19 MatrizOrdenada ← CriaMatriz(NumLinhas, NumColunas)
20 Para j = 1 até NumColunas
21     menorTupla ← MinHeap.Pop()
22     coluna ← tupla[3]
23     para i = 1 até NumLinhas
24         MatrizOrdenada[i, j] ← Matriz[i, coluna]
25     fim
26 fim
27 devolva MatrizOrdenada ▷ Matriz ordenada que facilita a visualização da cascata
28 fim

```

Referências

- [BORNHOLDT 2008] Stefan BORNHOLDT. “Boolean network models of cellular regulation: prospects and limitations”. *Journal of The Royal Society Interface* 5.suppl_1 (2008), S85–S94. DOI: [10.1098/rsif.2008.0132.focus](https://doi.org/10.1098/rsif.2008.0132.focus). eprint: <https://royalsocietypublishing.org/doi/pdf/10.1098/rsif.2008.0132.focus>. URL: <https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2008.0132.focus> (citado na pg. 23).
- [DAVIDICH e BORNHOLDT 2013] Maria I. DAVIDICH e Stefan BORNHOLDT. “Boolean network model predicts knockout mutant phenotypes of fission yeast”. *PLOS ONE* 8.9 (set. de 2013), pp. 1–9. DOI: [10.1371/journal.pone.0071786](https://doi.org/10.1371/journal.pone.0071786). URL: <https://doi.org/10.1371/journal.pone.0071786> (citado nas pgs. 5, 6).
- [LEE e HUANG 2009] Wen-Bin LEE e Jiun-Yan HUANG. “Robustness and topology of the yeast cell cycle boolean network”. *FEBS Letters* 583.5 (2009), pp. 927–932. ISSN: 0014-5793. DOI: <https://doi.org/10.1016/j.febslet.2009.02.010>. URL: <https://www.sciencedirect.com/science/article/pii/S0014579309001082> (citado na pg. 25).
- [LI *et al.* 2004] Fangting LI, Tao LONG, Ying LU, Qi OUYANG e Chao TANG. “The yeast cell-cycle network is robustly designed”. *Proceedings of the National Academy of Sciences* 101.14 (2004), pp. 4781–4786. DOI: [10.1073/pnas.0305937101](https://doi.org/10.1073/pnas.0305937101). eprint: <https://www.pnas.org/doi/pdf/10.1073/pnas.0305937101>. URL: <https://www.pnas.org/doi/abs/10.1073/pnas.0305937101> (citado nas pgs. 3, 4, 7, 8, 23).
- [THIEFFRY 2007] Denis THIEFFRY. “Dynamical roles of biological regulatory circuits1”. *Briefings in Bioinformatics* 8.4 (jul. de 2007), pp. 220–225. ISSN: 1467-5463. DOI: [10.1093/bib/bbm028](https://doi.org/10.1093/bib/bbm028). eprint: <https://academic.oup.com/bib/article-pdf/8/4/220/591211/bbm028.pdf>. URL: <https://doi.org/10.1093/bib/bbm028> (citado na pg. 23).
- [TRAN *et al.* 2013] Van TRAN, Matthew McCALL, Helene McMURRAY e Anthony ALMUDEVAR. “On the underlying assumptions of threshold boolean networks as a model for genetic regulatory network behavior”. *Frontiers in Genetics* 4 (2013). ISSN: 1664-8021. DOI: [10.3389/fgene.2013.00263](https://doi.org/10.3389/fgene.2013.00263). URL: <https://www.frontiersin.org/articles/10.3389/fgene.2013.00263> (citado na pg. 12).

