
Projeto de Pesquisa

Iniciação Científica

Caracterização Topológica
de Dinâmica de Redes Biológicas
Modeladas por Redes Booleanas

Aluno: Miguel Pereira Ostrowski
Prof. Dr. Ronaldo Fumio Hashimoto
ronaldo@ime.usp.br

Resumo

Muitas das análises da dinâmica de rede Booleanas são feitas pela caracterização de bacias de atração (seus tamanhos, por exemplo) e de seus atratores (quantidade de pontos fixos, por exemplo). Apesar dessas análises, outras características poderiam ser exploradas. Uma delas, é o caminho central. Neste projeto, temos como objetivo geral, estudar que medidas topológicas caracterizam a dinâmica (o grafo do diagrama de estados) de redes biológicas modeladas por Redes Booleanas. Posteriormente, podemos ainda estudar quais propriedades matemáticas da rede de interação que produzem tais diagramas de estados.

Palavras Chaves: Biologia de Sistemas, Rede Booleana, Rede Complexa.

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*), introduzido na biologia por [3]. Neste modelo, as expressões de seus componentes são representadas por apenas dois valores: “ligado” ou “desligado”.

Neste projeto, uma rede Booleana $G(V, C)$ (também conhecida como rede Booleana com limiar) é formada por um conjunto de nós (ou genes) $V = \{x_1, x_2, \dots, x_n\}$ e uma *rede de interação* (ou *matriz de conectividade*) $A_{n \times n}$ definida como

$$a_{i,j} = \begin{cases} -1, & \text{para uma regulação negativa de } x_j \text{ para } x_i \\ 1, & \text{para uma regulação positiva de } x_j \text{ para } x_i \\ 0, & \text{caso contrário.} \end{cases} \quad (1)$$

Veja um exemplo na Fig. 1.

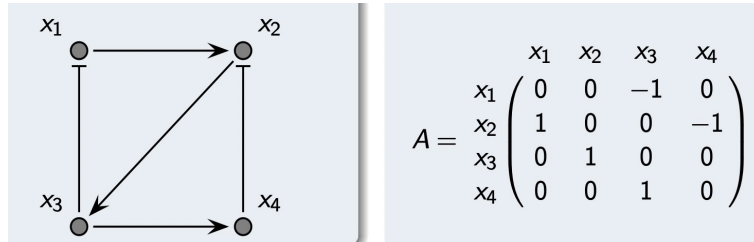


Figura 1: Exemplo de uma rede de interação e sua representação matricial.

O *signal de entrada* $\mathcal{I}_i(t)$ de um gene x_i no tempo t depende dos estados “ligado” ou “desligado” dos genes em V no mesmo tempo t e da linha i da

matriz de conectividade $A_{n \times n}$ da seguinte forma:

$$\mathcal{I}_i(t) = \sum_{j=1}^n a_{i,j} \cdot x_j(t). \quad (2)$$

O estado do gene x_i no tempo $t + 1$ depende de seu sinal de entrada $\mathcal{I}_i(t)$ da seguinte forma:

$$x_i(t+1) = \begin{cases} 1, & \text{se } \mathcal{I}_i(t) > 0 \\ 0, & \text{se } \mathcal{I}_i(t) < 0 \\ x_i(t), & \text{se } \mathcal{I}_i(t) = 0. \end{cases} \quad (3)$$

Como cada gene $x_i(t)$ no tempo t pode assumir somente 2 valores, o vetor $s(t) = [x_1(t), \dots, x_n(t)]$ 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 $s(t) \in \{0, 1\}^n$. 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 $s(t + 1)$. Podemos representar estas transições de estado através de um *diagrama de transição de estados*.

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, etc ...

2 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 projeto, 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 seqüência 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 levedura, Li et al. [4] 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. 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. [4] 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 proposta por Li et al. [4] é mostrada na Fig. 2. A partir desta figura, é possível construir uma matriz de regulação $A_{11 \times 11}$ da seguinte forma: (i) uma flecha de um gene i para um gene j representa “ativação” do gene j pelo gene i , e neste caso, $a_{ji} = 1$; (ii) uma linha de um gene i para um gene j com uma barra no final representa “repressão” do gene j pelo gene i , e neste caso $a_{ji} = -1$; (iii) laços pontilhados representam auto-degradação, e neste caso, $a_{ii} = -0.1$; (iv) para os casos restantes, $a_{ji} = 0$.

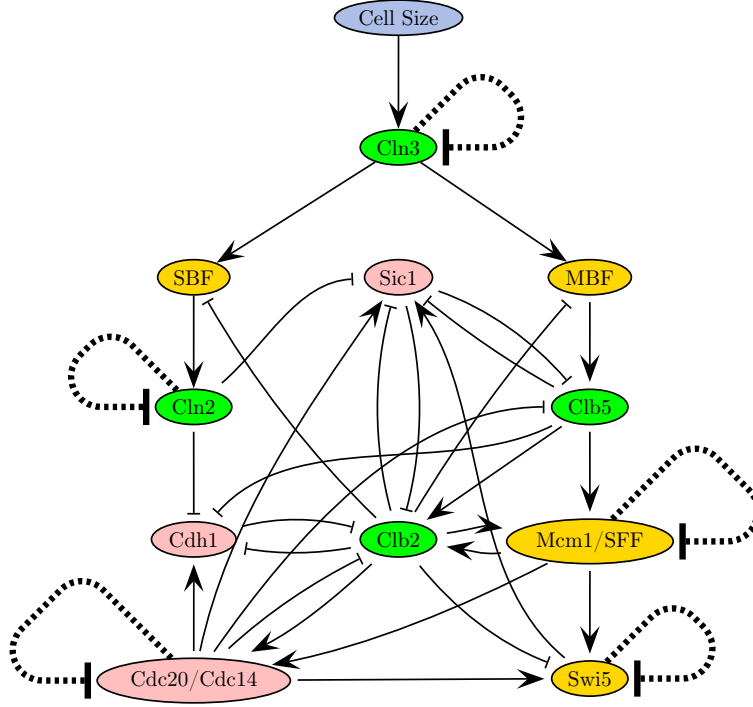


Figura 2: Rede de interações proposta por Li et al [4].

A dinâmica da rede do Li et al. [4] 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

A bacia de atração de maior tamanho tem aproximadamente 80% dos estados possíveis e está representada na Fig. 3. Nesta bacia há uma *trajetória central* com 13 estados (marcada pelos arcos em azul) que representa o ciclo celular da levedura (modela um fenômeno biológico) e por isso consideramos esta ser esta rede Booleana uma *rede biológica* de acordo com a 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

Observe que cada estado na trajetória corresponde a alguma fase da divisão celular da levedura.

3 Caminho Central

Muitas das análises da dinâmica da rede são feitas pela caracterização de bacias de atração (seus tamanhos, por exemplo) e de seus atratores (quantidade de pontos fixos, por exemplo). Apesar dessas análises, outras características poderiam ser exploradas. Uma delas, é o caminho central.

Nem todas as redes Booleanas contêm uma trajetória central tão evidente na maior bacia de atração como mostrada na Fig. 3. De fato, podemos ter diferentes topologias como mostradas na Fig. 4. Enquanto temos uma topologia “estrela” na Fig. 4(a), a topologia apresentada na Fig. 4(b) tem uma característica mais próxima da rede da levedura apresentada em [4] e portanto sendo mais biológica. Dessa forma, uma pergunta científica interessante que medidas topológicas (como por exemplo, as apresentadas em [2]) aplicadas à maior bacia de atração poderiam caracterizar redes biológicas, e vice-versa.

4 Objetivos

Como objetivo geral, este projeto estudará que medidas topológicas caracterizam a dinâmica (o grafo do diagrama de estados) de redes biológicas modeladas por Redes Booleanas. Posteriormente, podemos ainda estudar quais propriedades matemáticas da rede de interação que produzem tais diagramas de estados.

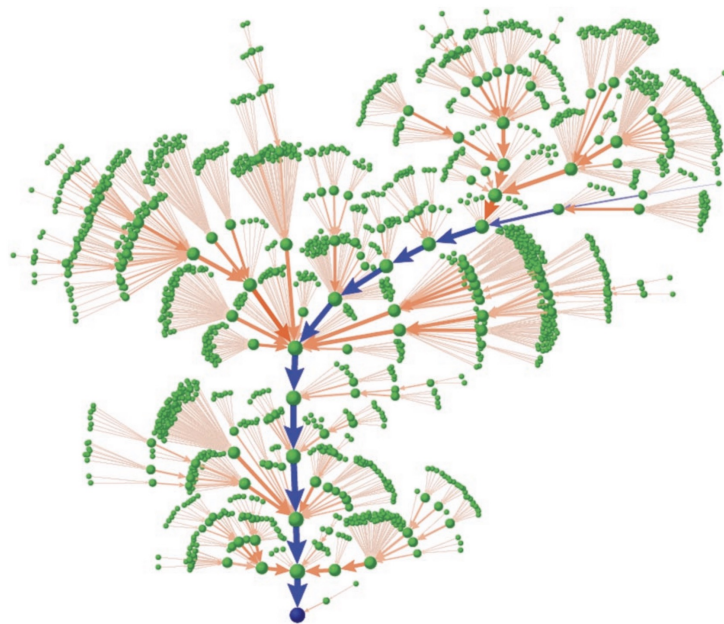


Figura 3: As trajetórias dos 1.764 estados (nós verdes) fluindo para o ponto fixo G_1 (nó azul). Os arcos entre os estados indicam a direção de um estado para outro. A sequência do ciclo celular é indicada pelos arcos em azul. O tamanho de um nó e a espessura de um arco são proporcionais ao logaritmo do número de trajetórias que passa por eles. Figura extraída de [4].

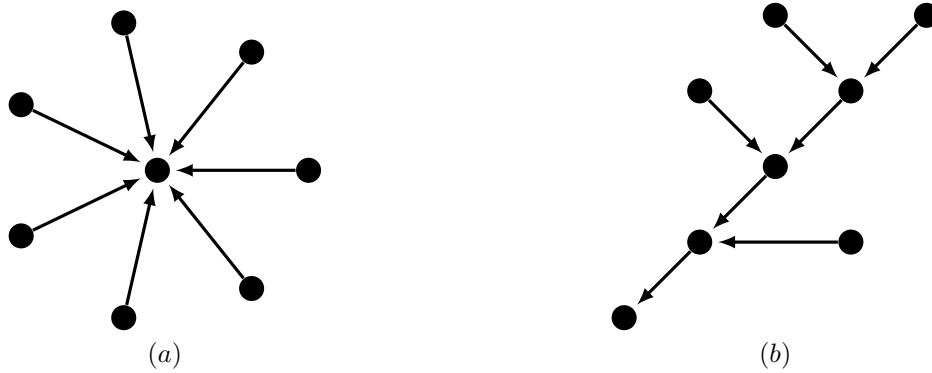


Figura 4: Diferentes topologias para o diagrama de estados.

5 Metodologia

Uma forma de atingir os objetivos deste projeto, propomos uma metodologia dividida em 3 camadas.

Na primeira camada, a proposta é estudar e aplicar medidas topológicas [2] em grafos diagramas de estados. Podemos inicialmente aplicar medidas topológicas já apresentadas na literatura e comumente utilizadas tais como: assortatividade (ASS), grau médio (DEG), grau máximo (MAX), grau mínimo (MIN), centralidade de intermediação média (BET), coeficiente de agrupamento (CC), média comprimento do caminho curto (ASPL), desvio padrão médio (SD), frequência de motivos com tamanho 3 (MT3) e frequência de motivos com tamanho 4 (MT4) [1, 2].

Estas medidas poderiam ser aplicadas em técnicas de seleção de características utilizadas em Aprendizagem de Máquina (por exemplo, utilizando árvores de decisão) para separar diagramas de estados de redes biológicas (como por exemplo, a rede apresentada em [4]) e de redes que têm dinâmicas que claramente não são biológicas.

Em uma segunda camada, podemos utilizar medidas que são próprias de diagramas de estados. Por exemplo, a quantificação da “convergência” de trajetórias de uma rede proposta em [4], denotada por W e definida da seguinte forma. Vamos denotar por $T_{j,k}$ o número total de trajetórias a partir de todos os estados da rede que passam do estado j para o estado k (veja os pesos dos arcos da Fig. 5). Agora, para cada estado s da rede, podemos definir seu peso como uma medida da convergência de sua trajetória ao atrator. De fato, se a trajetória do estado s ao seu atrator tem L transições, digamos $(s = s_0, s_1, \dots, s_L)$, então a convergência w_s é definida como

$$w_s = \frac{1}{L} \sum_{k=1}^L T_{i-1,i}.$$

A sobreposição de todas trajetórias de uma rede, denotada por W , é definida

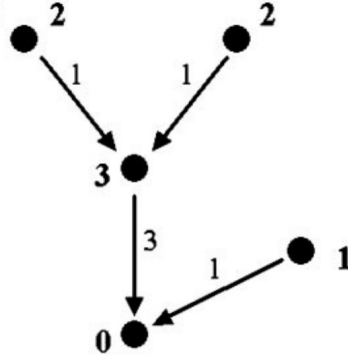


Figura 5: Ilustração esquemática da definição de w_s . O peso de cada arco indica o número total de trajetórias que passam pelo arco. O peso de cada nó s é a medida da convergência de sua trajetória ao atrator. A média da convergência é $W = 8/5 = 1.6$. Figura extraída de [4].

como a média das convergências das trajetórias a partir de todos estados.

E para finalizar, em uma última camada, podemos explorar e propor novas medidas topológicas que poderiam ser úteis para atingir os objetivos do projeto.

6 Cronograma de Execução

Apresentamos abaixo um cronograma de execução (dividido em doze meses) do projeto.

Atividade	01	02	03	04	05	06	07	08	09	10	11	12
Estudar medidas topológicas conhecidas	XX	XX										
Estudar medidas para diagrama de estados		XX	XX	XX								
Propor e explorar novas medidas topológicas			XX	XX	XX							
Estudar métodos de seleção de características					XX	XX	XX					
Implementação		XX		XX		XX	XX	XX				
Teste e Análise de Resultados		XX		XX		XX	XX	XX	XX	XX		
Escrita de um relatório										XX	XX	
Apresentação dos resultados no SIICUSP												XX

Referências

- [1] S Boccaletti, V Latora, Y Moreno, M Chavez, and D Hwang. Complex networks: Structure and dynamics. *Physics Reports*, 424(4-5):175–308, February 2006.
- [2] L. da F. Costa, F. A. Rodrigues, G. Travieso, and P. R. Villas Boas. Characterization of complex networks: A survey of measurements. *Adv. Phys.*, 56(1):167–242, 2007.
- [3] S. A. Kauffman. Metabolic Stability and Epigenesis in Randomly Constructed Genetic Nets. *Journal of Theoretical Biology*, 22:437–467, 1969.
- [4] F. Li, T. Long, Y. Lu, Q. Ouyang, and C. Tang. The Yeast Cell-Cycle Network is Robustly Designed. *Proceedings of the National Academy of Sciences of the United States of America*, 101(14):4781–4786, 2004.