

MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO RIO GRANDE
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM COMPUTACIONAL

**DESENVOLVIMENTO DE UM FRAMEWORK PARA A MODELAGEM E
A SIMULAÇÃO DE REDES REGULATÓRIAS GENÉTICAS USANDO
SISTEMA MULTIAGENTE (FMAS-GENERE)**

por

Nilzair Barreto Agostinho

Banca

Prof.^a Dr.^a Diana Francisca Adamatti (Orientadora - FURG)

Prof.^o Dr.^o Adriano V. Werhli (Co-Orientador - FURG)

Prof.^o Dr.^o Leonardo Emmendorfer (Membro do PPGMC - FURG)

Prof.^a Dr.^a Karina dos Santos Machado (Membro Externo - FURG)

Prof.^o Dr.^o Marilton Sanchotene de Aguiar (Membro Externo - UFPel)

Prof.^a Dr.^a Mariana Recamonde Mendoza (Membro Externo - UFRGS)

Tese de Doutorado em Modelagem Computacional

Rio Grande, 27 de julho de 2022.

Nilzair Barreto Agostinho

**“Desenvolvimento de uma Ferramenta para a Modelagem e a Simulação de Redes
Regulatórias Genéticas usando Sistema Multiagente”**

Tese apresentada ao Programa de Pós-Graduação em Modelagem Computacional da Universidade Federal do Rio Grande - FURG, como requisito parcial para obtenção do Grau de Doutora. Área de concentração: Modelagem Computacional.

Aprovada em:

BANCA EXAMINADORA

Documento assinado digitalmente
 Diana Francisca Adamatti
Data: 25/07/2022 10:41:05-0300
Verifique em <https://verificador.iti.br>

Profa. Dra. Diana Francisca Adamatti (Orientadora – FURG)

Documento assinado digitalmente
 KARINA DOS SANTOS MACHADO
Data: 28/07/2022 23:07:33-0300
Verifique em <https://verificador.iti.br>

Profa. Dra. Karina dos Santos Machado (FURG)

Documento assinado digitalmente
 Leonardo Ramos Emmendorfer
Data: 29/07/2022 10:17:04-0300
Verifique em <https://verificador.iti.br>

Prof. Dr. Leonardo Ramos Emmendorfer (UFSM)

Documento assinado digitalmente
 MARIANA RECAMONDE MENDOZA GUERREIRO
Data: 27/07/2022 21:03:58-0300
Verifique em <https://verificador.iti.br>

Profa. Dra. Mariana Recamonde Mendoza (UFRGS)

Documento assinado digitalmente
 MARILTON SANCHOTENE DE AGUIAR
Data: 28/07/2022 08:51:49-0300
Verifique em <https://verificador.iti.br>

Prof. Dr. Marilton Sanchotene de Aguiar (UFPel)

Rio Grande - RS
2022

Ficha Catalográfica

A275d Agostinho, Nilzair Barreto.

Desenvolvimento de um *Framework* para a modelagem e a simulação de redes regulatórias genéticas usando sistema multiagente (FMAS-GENERE) / Nilzair Barreto Agostinho. – 2022. 96 f.

Tese (doutorado) – Universidade Federal do Rio Grande – FURG, Programa de Pós-Graduação em Modelagem Computacional, Rio Grande/RS, 2022.

Orientadora: Dra. Diana Francisca Adamatti.

Coorientador: Dr. Adriano Velasque Werhli.

1. Sistemas Multiagente 2. Redes Regulatorias Genéticas
I. Adamatti, Diana Francisca II. Werhli, Adriano Velasque III. Título.

CDU 004.43

Catálogo na Fonte: Bibliotecário José Paulo dos Santos CRB 10/2344

MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO RIO GRANDE
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM COMPUTACIONAL
FURG

Nilzair Barreto Agostinho

Desenvolvimento de um Framework para a Modelagem e a Simulação de Redes Regulatórias Genéticas usando Sistema Multiagente (FMAS-GENERE)

DOUTORADO EM MODELAGEM COMPUTACIONAL

Tese apresentada à Banca Examinadora como
exigência parcial para obtenção do título de Dou-
tora em Modelagem Computacional pela Univer-
sidade Federal do Rio Grande, sob a orientação da
Prof.^a Doutora Diana Francisca Adamatti

Rio Grande - RS
2022

AGRADECIMENTOS

Ao Senhor meu Redentor que me guia, orienta, me levanta, me inspira, a quem dou graças em tudo.

À minha família em especial ao meu esposo, minha filhinha, meus pais e meu irmão por me apoiarem, me ouvirem, me darem todo suporte, amor e carinho e por compreenderem todos os momentos que precisei estar ausente, por me ajudarem quando eu não podia mais me levantar, quando eu “surtava” e quando eu estava doente e não podia fazer praticamente nada, literalmente.

À minha orientadora Diana Francisca Adamatti por me ensinar, orientar, guiar, me “levantar” e nunca me deixar desistir em momentos de dificuldade. Ao meu co-orientador Adriano Velasques Werhli por seus ensinamentos, orientação e por me guiar no “mundo da Bioinformática” desde os tempos de Graduação.

Aos amigos novos e antigos por me apoiarem, entenderem minha ausência e estarem ao “meu lado” mesmo distantes, alguns muito distantes geograficamente. Em especial minha amiga Sylvia que esteve presente nos piores e nos melhores momentos, aos amigos Dania, Roberta, Rossana, Juliana e Victor. Ao meu amigo Dino que compreendeu anos de minha ausência e mesmo assim seu amor continua o mesmo. Ao meu amigo Karl Bischoff que foi meu amparo em muitos momentos, que me ajudou a superar muitos momentos ruins, que foi meu professor de inglês e esteve sempre presente mesmo muito distante geograficamente. E por fim, mas não menos importante aos amigos Vinicius Martins e Jean Blay que estiveram presentes para uma boa conversa, para troca de conhecimentos, para me ouvir, para indicar músicas novas para a playlist para me ajudar a concentrar e que sempre me fazem sorrir. E à minha amiga Daiane pelo chimarrão dos horários de almoço, das noites quentes e geladas, pelas conversas, conselhos e risos que faziam meu dia ensolarado mesmo quando estava nublado.

Em meu trabalho (Compass), aos colegas e aos gestores que me apoiaram e ajudaram a crescer pessoal e profissionalmente durante esse período de dupla jornada.

À FURG que me proporcionou todos esses anos de ensino, aprendizado e experiência desde meu período de Graduação, aos professores que fizeram parte dessa jornada e aos técnicos e funcionários ao longo de todos esses anos.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo período de concessão de bolsa de estudos.

RESUMO

Os sistemas biológicos são altamente complexos e a sua separação em partes individuais torna tratável seu estudo. A representação de sistemas biológicos como Redes Regulatórias Genéticas (Genetic Regulatory Networks - GRN) que formam um mapa das interações entre as moléculas num organismo é uma maneira padrão de representar essa complexidade biológica. As GRN são compostas de genes que podem ser apenas modulados por outros genes ou traduzidos em fatores de transcrição, que por sua vez regulam outros genes. Os cientistas trabalharam na inferência e representação de GRNs. Para fins de simulação e inferência, muitos modelos matemáticos e algorítmicos diferentes foram adotados para representar as GRN nos últimos anos. Entre esses métodos, é formulada a hipótese de que os Sistemas Multiagente (MAS - Multi-Agent System) foram pouco explorados. Neste trabalho, apresenta-se os esforços, trabalhos e resultados no desenvolvimento de uma ferramenta para modelagem e simulação de GRN usando MAS. Para tal, é proposto um MAS composto por agentes que imitam os processos bioquímicos de regulação de genes. A ferramenta foi desenvolvida em etapas, de forma que seu desenvolvimento foi evoluindo e cada uma das versões foi validada de forma diferente, até ter o modelo final. Primeiramente, foi validado através da comparação com os resultados da teoria de Michaelis e Menten. Em seguida, através da comparação com Modelo de Uri Alonn (baseado em motifs). Finalmente, a ferramenta foi concluída e validada através do estudo de caso da planta *Arabidopsis Thaliana*. Esta versão mostra-se flexível para realizar simulações de GRNs de acordo com as configurações estabelecidas nos arquivos de entrada e assim oferecer a autonomia e a proatividade característica dos MAS ao simular GRNs genéricas. Sendo assim, apesar de serem usados estudos de casos para validação de cada uma das etapas de desenvolvimento, o modelo desenvolvido é capaz de realizar simulações de GRNs diversas e genéricas, por meio da configuração/parametrização dos arquivos de entrada na forma das estruturas de motifs.

Palavras-chaves: Sistemas Multiagente, Redes Regulatórias Genéticas

ABSTRACT

Biological systems are highly complex and their separation into parts makes their study tractable. The representation of biological systems as genetic patterns (Genetic Regulatory Network) that form a map of biological ways as representative patterns as a set is a biological variety. As GRN they are composed of genes that can only be modulated or translated into other composting factors, which in turn regulate other genes. Scientists worked on the inference and representation of GRNs. For simulation and inference purposes, many different mathematical and algorithmic models have been adopted to represent as GRN in recent years. Among these methods, a hypothesis is formulated that Multi-Agent Systems (MAS - Multi-Agent System) were few explorers. In this work, we present the efforts, works and results in the development of a tool to model and simulate GRN using MAS. In this way, a MAS composed of agents that mimic the biochemical processes of gene regulation is proposed. The tool was developed in stages, so that its development evolved and each version was validated differently for the final model. Firstly, it was validated through the comparison with the results of the theory of Michaelis and Menten. After, the comparison was with the Uri model (based on a single reason). Finally, the tool was concluded and validated through the Arabidopsis study. The final version allows GRN simulations according to the settings in the input files, in this way, there are MAS autonomy and pro-activity characteristics. Therefore, even using case studies for validation of each one of the development stages, the model developed is capable of performing simulations of diverse and general GRNs through configuration/parameterization of the input files, based on motif structures.

Keywords: Multiagents Systems, Genetic Regulatory Networks

ÍNDICE

1	Introdução	10
1.1	Objetivos do Trabalho	11
1.2	Estrutura do Texto	12
2	Referencial Teórico e Aplicado	13
2.1	Redes Regulatórias Genéticas	13
2.2	Métodos para representação de Redes Regulatórias Genéticas	14
2.2.1	Redes Booleanas	14
2.2.2	Redes Bayesianas	14
2.2.3	Redes Bayesianas Dinâmicas	15
2.2.4	Inferência de Redes Regulatórias com amostrador MCMC guiado por Informações Mútuas	15
2.2.5	ODE	16
2.2.6	Motifs	16
2.3	Sistemas Multiagente	17
2.3.1	NetLogo	18
2.4	Base teórica para os Estudos de Caso	23
2.4.1	Processo de Michaelis-Menten	23
2.4.2	Uri Alon	24
2.4.3	Ciclo circadiano Eucariótico da Planta Arabidopsis Thaliana	25
3	Revisão Sistemática	28
3.1	Descrição dos trabalhos relacionados resultantes da SLR	39
3.2	Revisão Sistemática Complementar	42
4	Metodologia	47
5	Estudos de caso	54
5.1	Comparação com os resultados da teoria de Michaelis e Menten	54
5.1.1	Resultados	54
5.2	Comparação com o Modelo de Uri Alon	57
5.2.1	Metodologia	57
5.2.2	Resultados	59
5.3	Comparação com o Modelo da planta Arabidopsis Thaliana	71
5.3.1	Metodologia	71
5.3.2	Resultados	75
5.3.3	Comparação do Modelo FMAS-GENERE com o modelo BioASF-BIOPAX	85
6	Conclusões e Trabalhos Futuros	90

LISTA DE FIGURAS

1.1	Dogma Central da Biologia	10
2.1	Exemplo de rede Bayesiana	15
2.2	Biblioteca do Netlogo	19
2.3	Modelo “Virus” antes de inicializar a simulação	19
2.4	Modelo “Virus” com a simulação em andamento	20
2.5	Aba informação no Netlogo	21
2.6	Aba código no Netlogo	22
2.7	Abordagem de Michaelis-Menten	24
2.8	Motifs	25
2.9	Modelo Arabidopsis Thaliana	26
2.10	Modelo Arabidopsis Thaliana detalhado	27
3.1	SLR Detalhado	29
3.2	SLR Detalhado - continuação	30
3.3	Modelo em espiral	33
3.4	Adaptada de [Haydarlou et al., 2016]: O modelo tem como entrada o modelo de rede biológica no BioPAX, parâmetros físicos iniciais, regras da simulação e regras de análise. A saída é um sistema multiagente que é gerado automaticamente por BioASF e é responsável pela execução de interações BioPAX	40
3.5	Comportamento da simulação utilizando uma porta AND	42
4.1	Interface do modelo Inicial	49
4.2	Motifs x representação do Sítio de ligação e TFs	51
4.3	Funcionamento do Modelo - esquema de funcionamento do framework	52
4.4	Escala que representa os períodos definidos no modelo.	53
5.1	Modelo Inicial - 126 ticks	55
5.2	Modelo Inicial - 1042 ticks	55
5.3	Modelo Inicial - 1386 ticks	56
5.4	Modelo Inicial - 1944 ticks	56
5.5	Resultados Comparação Uri Alon - Produto mediante o funcionamento do EF e o funcionamento da porta AND	62
5.6	Resultados comparação Uri Alon - Produto mediante o funcionamento do EF e o funcionamento da porta AND (Inibição)	65
5.7	Resultados EF - portas AND e OR	69
5.8	Gráfico - comportamento Modelo Uri Alon	70
5.9	Concentração do gene LHYmod produzida	76
5.10	Concentração do gene PRR9 produzida	77
5.11	Concentração do gene PRR7 produzida	78
5.12	Concentração do gene NI produzida	79

5.13	Concentração do gene GI produzida	80
5.14	Concentração do gene TOC1mod produzida	81
5.15	Concentração do gene Y produzida	82
5.16	Concentração do gene LHY produzida	83
5.17	Concentração do gene ZTL produzida	84
5.18	Concentrações dos genes PRR9, PRR7, NI, LHY e TOC1 resultantes da simulação realizada pela ferramenta BioPAX adaptado de [Pokhilko et al., 2010]	86
5.19	Concentração dos genes PPR9, PRR7 e NI	88
5.20	Concentração dos genes LHY e TOC1	89

LISTA DE TABELAS

1	Protocolo de definição da SLR	31
2	Palavras Chaves	32
3	Resultados (resumido) após etapa de filtro por resumo	35
4	Resultado da SLR - Comparação entre artigos - Parte I	37
5	Resultado da SLR - Comparação entre artigos - Parte II	38
6	Bases de dados e todas as variantes de palavras-chave - PUBMED	43
7	Bases de dados e todas as variantes de palavras-chave - PMC	44
8	Bases de dados e todas as variantes de palavras-chave - Science Direct	45
9	Bases de dados e todas as variantes de palavras-chave - Scopus	46
10	Número de artigos em cada etapa	46
11	Definição dos agentes	47
12	Definição das restrições	47
13	Exemplo de arquivo de entrada de definições - Ativação (porta AND) - Comparação Uri Alon	57
14	Exemplo de arquivo de entrada de restrições - Ativação (porta AND) - Comparação Uri Alon	58
15	Arquivo de entrada de definições - Ativação (porta AND) - Resultados - Comparação Uri Alon	59
16	Arquivo de entrada de restrições - Ativação (porta AND) - Resultados - Comparação Uri Alon	60
17	Arquivo de entrada de definições - Ativação (porta OR) - Resultados - Comparação Uri Alon	60
18	Arquivo de entrada de restrições - Ativação (porta OR) - Resultados - Comparação Uri Alon	60
19	Arquivo de entrada de definições - Inibição (porta AND) - Resultados - Comparação Uri Alon - 2	63
20	Arquivo de entrada de restrições - Inibição (porta AND) - Resultados - Comparação Uri Alon - 2	63
21	Arquivo de entrada de definições - Inibição (porta OR) - Resultados - Comparação Uri Alon - 2	63
22	Arquivo de entrada de restrições - Inibição (porta OR) - Resultados - Comparação Uri Alon - 2	64
23	Genes/Fatores de transcrição do ciclo da Arabidopsis Thaliana	71
24	Interação - Tradução para “motifs”	72
25	Recorte de um arquivo de entrada - Atributos	72
26	Recorte de um arquivo de entrada - Restrições	74

LISTA DE ABREVIATURAS

BN Bayesian networks - Redes Bayesianas

BoN Boolean networks - Redes Booleanas

CCA1 CIRCADIAN CLOCK ASSOCIATED 1

C1-FFL Um dos tipos de “motifs” - feedforward loop (“the coherent type-1 FFL (C1-FFL)”)

DNA Ácido Desoxirribonucleico

ED Equações Diferenciais

EF External Factor - Fator Externo

FFL Um dos tipos de “motifs” - feedforward loop (“FFL”)

GI GIGANTEA

GRN Genetic Regulatory Networks - Rede Regulatória Genética

LHY LATE ELONGATED HYPOCOTYL

LHYmod LATE ELONGATED HYPOCOTYL modified

MAS MultiAgent Systems - Sistemas Multiagente

ODE Ordinary differential Equation

PBN Probabilistic Boolean Networks - Redes Booleanas Probabilísticas

PRR5 PSEUDO-RESPONSE REGLATOR 5

PRR7 PSEUDO-RESPONSE REGLATOR 7

PRR9 PSEUDO-RESPONSE REGLATOR 9

RNA Ácido Ribonucleico

RNA^m RNA Mensageiro - Ácido Ribonucleico responsável pela transferência de informações do DNA até o citoplasma

SLR Systematic Literature Review - Revisão Sistemática

SMA Sistemas Multiagente

TF/gene Transcription Factor - Fator de Transcrição

TFs/genes Transcription Factors - Fatores de Transcrição

TOC1 TIMING OF CAB EXPRESSION 1

TOC1mod TIMING OF CAB EXPRESSION 1 modified

Y Hypothetical activator of TOC1

ZTL F-box ZEITLUPE

1 INTRODUÇÃO

Embora o dogma central da biologia afirme que a informação flui através de macro-moléculas, do DNA para o RNA e do RNA para as proteínas, conforme a a Figura 1.1, a vida não existiria apenas das macro-moléculas [Schreiber, 2005]. Então, para que o dogma central seja altamente descritivo, ele deve incluir pequenas moléculas. Essas pequenas moléculas são os elementos-chave em vários tópicos das ciências da vida como as origens da vida, memória e cognição, detecção e sinalização, compreensão do circuito celular e tratamentos de doenças [Schreiber, 2005].

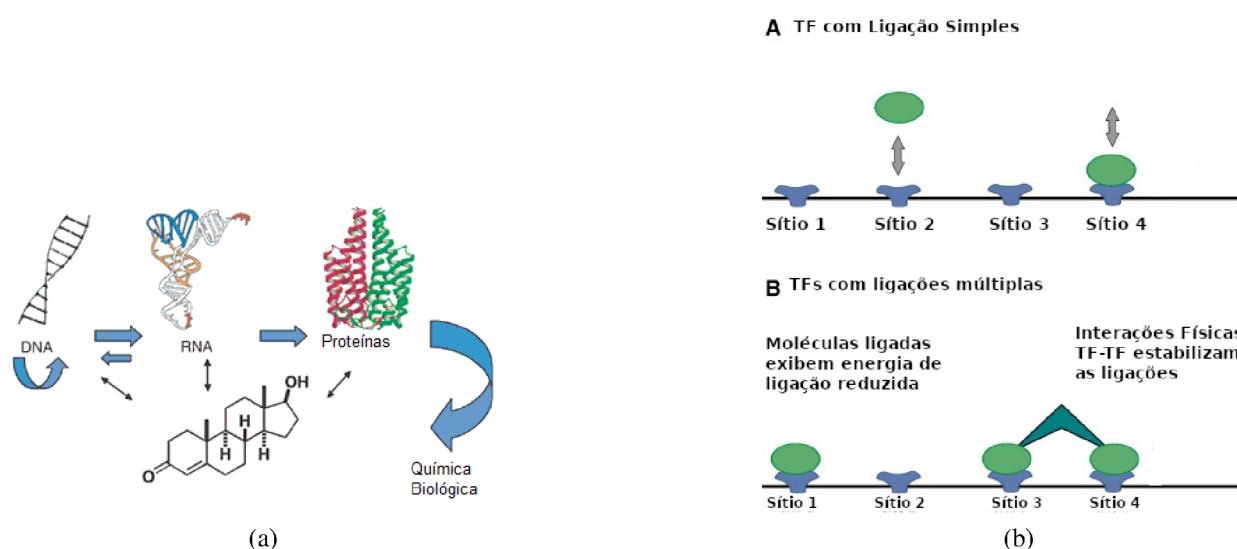


Figura 1.1: (a) Dogma Central da Biologia. Pode-se verificar nesta Figura passos da replicação, após a transcrição e, finalmente obtendo o produto final (proteína) - adaptado de [Schreiber, 2005]. (b) Fatores de Transcrição e Sítios de Ligação. Observa-se que os TF são proteínas que ajudam a ativar ou desativar genes específicos, ligando-se ao DNA próximo. Os Fatores de Transcrição, que são ativadores, aumentam a transcrição de um gene. Repressores diminuem a transcrição - adaptado de [Ezer, 2014].

Parte dos processos do dogma central, como mostrado na Figura 1.1a, pode ser representada como uma rede de genes que se regulam. Esses genes (pequenas moléculas) se traduzem em proteínas que podem ativar ou inibir a transcrição de outros genes. Portanto, identificar os padrões de Redes Regulatórias Genéticas (Genetic Regulatory Networks - GRN) é de grande valia para uma melhor compreensão dos processos que ocorrem em nível molecular. Então, no cenário atual da área de Sistemas Biológicos, trabalha-se para o desenvolvimento de ferramentas que possam auxiliar na obtenção de novos conhecimentos sobre as interações moleculares. Por este motivo, tem sido desenvolvidas pesquisas com o objetivo principal de obter um detalhamento da organização funcional de Sistemas Biológicos [Gibas and Jambeck, 2002, Lopes, 2011].

De acordo com estudos, a alta complexidade dos organismos está intimamente relacionada à organização de seus componentes em rede. Devido a essa complexidade, pesquisadores têm investido no estudo de redes biológicas [Lopes, 2011]. As GRN representam um mapeamento das interações que existem entre as moléculas de um organismo. Na representação de uma GRN, genes controlam a

expressão de outros genes. Porém, nem todos os genes da rede interagem entre si. Normalmente, um gene controla apenas um subconjunto da rede que, por sua vez, é controlado por outro subconjunto de genes [Lopes, 2011].

Métodos determinísticos não fornecem uma representação adequada dos componentes biológicos presentes no sistema em estudo, pois não fornecem resultados detalhados e minuciosos como o presente estudo necessita [Werlang et al., 2014]. Esta característica dificulta a inclusão de qualquer nova hipótese biológica para fins de teste diretamente no modelo. Normalmente, modelos determinísticos apresentam como resultados o comportamento médio de seus componentes. Já um Sistema Multiagente (Multiagent Systems - MAS) pode ser usado como um “framework” para modelagem e representação de redes regulatórias. Os MAS são ferramentas flexíveis que podem fazer a representação de cada indivíduo do sistema, facilitando a avaliação de novas hipóteses no modelo. MAS são conhecidos por terem poucos parâmetros para serem ajustados enquanto mantém sua capacidade de representar adequadamente um modelo biológico [Werlang et al., 2014].

Um exemplo de GRN são os ciclos circadianos, que são influenciados pela rotação da Terra ao redor do seu próprio eixo, o qual determina o ciclo de claro/escuro que se estende por 24h. Esse ambiente constante impôs uma pressão seletiva nos organismos, assim a maioria desenvolveu atividades comportamentais e metabólicas com periodicidade circadiana [Guerriero et al., 2012, Hearn et al., 2018]. Os ciclos circadianos estão presentes na maioria dos organismos e os ajudam a se adaptarem 24h dia/noite. Esses ciclos são compostos por um pequeno número de genes que ajudam nessa adaptação dos organismos. Pretende-se, então, utilizar como estudo de caso, os dados da planta *Arabidopsis Thaliana*. Esta planta é estudo de caso de muitos trabalhos e é utilizada como base no trabalho de modelagem determinística de [Pokhilko et al., 2012] e no trabalho de modelagem mista (determinística/estocástica) de [Guerriero et al., 2012].

Estudos mostram que foram encontradas diferenças importantes entre os modelos estocásticos discretos e sua aproximação determinística contínua e, na maioria dos casos, a estocasticidade mostrou ter um efeito positivo e robusto [Guerriero et al., 2012].

Assim, este trabalho mostra o desenvolvimento de um ambiente de simulação multiagente para vias regulatórias genéticas que necessita de poucos parâmetros para o funcionamento e possui a capacidade de representação de agente a agente.

1.1 Objetivos do Trabalho

Este trabalho tem como objetivo geral o desenvolvimento de um simulador multiagente para GRN. O modelo proposto é uma abstração dos processos bioquímicos subjacentes e simula a interação entre o TF (*Transcription Factor*) e um gene alvo.

As ferramentas MAS são muito flexíveis permitindo a representação de cada indivíduo em um sistema, simplificando a avaliação de novas hipóteses sobre o modelo. Assim, este trabalho apresenta um MAS como framework de modelagem para representação de uma rede regulatória.

Este trabalho tem como objetivos específicos:

- Modelagem e testes iniciais, utilizando Motifs;
- Modelagem e testes com as propostas de [Alon, 2007] e [Haydarlou et al., 2016];
- Validação dos dados resultantes, para realização de comparação dos resultados obtidos com resultados de modelos com abordagens diferentes, tais como as abordagens encontradas em [Alon, 2007] e [Haydarlou et al., 2016].

Esse simulador foi implementado na ferramenta NetLogo¹. Espera-se contribuir para melhores resultados na obtenção de valores e interações nas redes regulatórias.

1.2 Estrutura do Texto

O trabalho está organizado da seguinte maneira: o capítulo 2 apresenta o Referencial Teórico e Aplicado; no capítulo 3 é apresentada a Revisão Sistemática; o capítulo 4 apresenta a Metodologia; o capítulo 5 descreve os Estudos de Caso; e por último o capítulo 6 contém as Conclusões e trabalhos futuros da pesquisa.

¹<https://ccl.northwestern.edu/netlogo/>

2 REFERENCIAL TEÓRICO E APLICADO

Neste capítulo é apresentado o referencial teórico relacionado à GRN, aos MAS, bem como à ferramenta utilizada para o desenvolvimento do modelo, e à revisão sistemática realizada na busca de trabalho correlatos.

2.1 Redes Regulatórias Genéticas

As Redes Regulatórias Genéticas têm sido utilizadas como modelo para dispor mapas detalhados sobre as interações entre os genes e consequentemente sobre as interações moleculares e a organização funcional de organismos complexos. Porém, as GRN são altamente complexas e o processo para inferi-las é computacionalmente custoso [Lopes, 2011].

As Redes Regulatórias Genéticas (GRN) são um conjunto de genes que interagem uns com os outros num organismo. Tanto que alguns genes controlam a expressão de outros genes. Porém, nem todos os genes da rede interagem entre si. Normalmente, um gene controla apenas um subconjunto da rede e por sua vez é controlado por outro subconjunto de genes [Lopes, 2011].

As GRN são formadas por pequenas partículas, denominadas genes produtos (proteínas resultantes do processo que ocorre no Sítio de Ligação), que são ativados ou inibidos por proteínas chamadas Fatores de Transcrição (Transcription Factors - TF). Uma GRN descreve a interação entre os TFs e os genes que eles regulam. Os TFs podem responder a sinais biológicos e impactar a taxa de transcrição genética, ou seja, a produção do mRNA que mais tarde se traduz em proteína. Esse processo permite que as células produzam as proteínas necessárias nos momentos e quantidades adequados. Um exemplo da operação dos TFs é mostrado na Figura 1.1b.

Para que ocorra a produção de uma nova proteína, é preciso que a quantidade de TFs relacionados ao Sítio de Ligação do gene seja suficiente. Assim, os TFs se relacionam de acordo com o processo bioquímico e com fatores externos, de forma que possa ocorrer a produção de uma nova proteína ou a inibição de uma já existente.

Alguns dos métodos que têm sido utilizados para realizar a modelagem e/ou inferência de GRN são: Equações Diferenciais (ED) [Werhli, 2007], Redes Bayesianas (Bayesian Networks - BN) [Friedman et al., 2000], Redes Booleanas, (Boolean networks - BoN) [Shawn Martin et al., 2005], Redes Bayesianas Dinâmicas (DBN) [Werhli, 2007], Redes Booleanas Probabilísticas (Probabilistic Boolean Networks - PBN) [Shmulevich et al., 2002], Redes Bayesianas utilizando Markov Chain Monte Carlo e Graphical Gaussian Models (MCMC) [Agostinho et al., 2015], Redes Bayesianas utilizando MCMC guiado por informações mútuas [Agostinho et al., 2017], entre outras. Alguns desses métodos são descritos a seguir.

2.2 Métodos para representação de Redes Regulatórias Genéticas

2.2.1 Redes Booleanas

Os frameworks de modelagem qualitativa mais comuns para GRNs são baseados em redes Booleanas, que são capazes de descrever fenômenos biológicos como oscilações, eventos multi-estacionários, correlações de longo alcance, estabilidade de comportamento [Barbuti et al., 2020].

Uma rede booleana consiste em um conjunto de variáveis booleanas (os nós da rede) cada uma associada a uma função booleana definida em um subconjunto das variáveis. No contexto dos GRNs, cada variável representa o estado de ativação de um gene, que pode ser 1 (ativo) ou 0 (inativo). Além disso, a função booleana associada a uma variável representa a influência de outros genes no estado de ativação do gene descrito por tal variável. Na prática, tal função será utilizada periodicamente para atualizar a variável booleana correspondente, permitindo simular a dinâmica do GRN [Barbuti et al., 2020].

2.2.2 Redes Bayesianas

As BNs (Redes Bayesianas) são um dos métodos utilizados para representação de GRN. Elas são modelos gráficos direcionados utilizados para representar relações probabilísticas entre várias entidades que se relacionam entre si. Uma BN é composta por uma estrutura de grafo $\mathcal{M}$, uma família de distribuições probabilísticas $\mathcal{F}$ e um conjunto de parâmetros q . Já $\mathcal{M}$ é composta por um conjunto de nós e arestas, onde os nós representam variáveis e as arestas representam as relações entre as variáveis. A estrutura $\mathcal{M}$ tem que ser um gráfico acíclico direcionado (DAG) ou (PDAG - Parcialmente DAG) [Werhli and Husmeier, 2007]. O conjunto de parâmetros q indicam a natureza das interações entre os nós e a intensidade dessas interações [Werhli, 2007].

O cálculo da probabilidade conjunta em uma BN é caracterizada por uma regra simples e única em termos de probabilidades condicionais mais simples [Werhli, 2007]. Percebe-se então que são modelos estocásticos que incorporam aleatoriedade ou incerteza.

A descrição da estrutura de uma rede gênica é feita por meio de redes Bayesianas e é modelada por um grafo acíclico dirigido (DAG) $\mathcal{M} = (V; E)$. Os vértices $v_i \in V$, $1 \leq i \leq n$, representam genes e correspondem a variáveis aleatórias X_i . Considerando que v_i é um gene, então X_i descreve o nível de expressão do gene v_i . Para cada X_i , uma distribuição de probabilidade condicional $\mathcal{P}(X_i | \text{preditores}(X_i))$ é definida, na qual $\text{preditores}(X_i)$ representam as variáveis que regulam diretamente o gene v_i da rede $\mathcal{M}$ [Lopes, 2011].

No exemplo demonstrado na Figura 2.1 tem-se a representação de uma rede Bayesiana composta por 10 nós e 11 arestas, a qual representa uma rede regulatória de 10 genes.

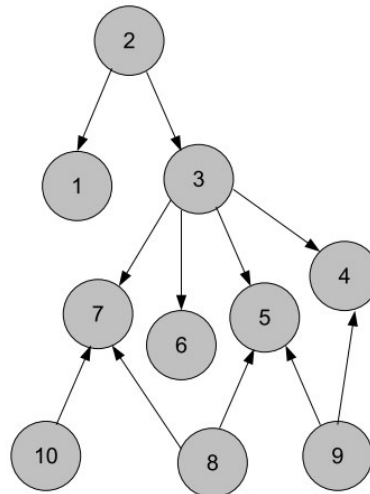


Figura 2.1: Exemplo de rede Bayesiana composta por 10 nós e 11 arestas

2.2.3 Redes Bayesianas Dinâmicas

As BNs mencionadas anteriormente têm algumas deficiências. Uma deficiência importante é que é impossível modelar loops de feedback, que são conhecidos por estarem presentes em redes biológicas reais. Também ao aplicar métodos MCMC (Markov Chain Monte Carlo) padrão é necessário verificar a aciclicidade das estruturas propostas; essa verificação de aciclicidade é um dos gargalos das simulações de MCMC. Uma maneira de resolver esses problemas é considerar Redes Bayesianas Dinâmicas (DBNs) [Werhli, 2007].

Considere-se que dois genes interagem entre si por meio de loops de retroalimentação. Observe-se que essa estrutura não é uma rede Bayesiana válida, pois viola a restrição de aciclicidade. Quando desdobra-se a rede no tempo, obtém-se um DAG adequado e, portanto, uma BN válida novamente, a chamada Rede Bayesiana Dinâmica (DBN) [Werhli, 2007].

2.2.4 Inferência de Redes Regulatórias com amostrador MCMC guiado por Informações Mútuas

A inferência computacionalmente eficiente e exata da topologia de rede regulatória é um problema aberto em Biologia de Sistemas. Neste tipo de representação é investigado o uso de informações prévias sobre a topologia da rede como guia para um amostrador de MCMC de estruturas de rede [Agostinho et al., 2015].

A informação prévia é obtida a partir de um método de inferência de rede mais grosseiro e rápido. Além disso, as redes regulatórias são representadas pelo modelo de Redes Bayesianas [Agostinho et al., 2015].

Os resultados mostram que o uso de informações prévias melhora drasticamente a convergência do amostrador MCMC. Portanto, o uso de métodos mais refinados se justifica, pois é provável que leve a resultados mais confiáveis com menos iterações de MCMC [Agostinho et al., 2015].

2.2.5 ODE

A GRN é um sistema complexo de alta dimensão, que pode ser representado por vários modelos matemáticos ou estatísticos. O modelo de equação diferencial ordinária (ODE) é um dos modelos GRN dinâmicos populares. Modelos de EDO lineares de alta dimensão têm sido propostos para identificar GRNs, mas com uma limitação da suposição do efeito de regulação linear [Wu et al., 2014].

Métodos como este precisam de especificações de equações diferenciais e suas variáveis e coeficientes, bem como suas calibrações e cálculos.

Como exemplo, o modelo proposto para a simulação da planta *Arabidopsis Thaliana* trata-se de um modelo ODE (Ordinary Differential Equation - Equação Diferencial Ordinária), que consiste em 28 equações diferenciais ordinárias e 104 parâmetros. Os valores de 43 parâmetros foram fixados com base nos dados disponíveis e 61 parâmetros foram ajustados a vários conjuntos de dados de séries temporais [Pokhilko et al., 2010].

Para a planta *Arabidopsis Thaliana* os valores dos parâmetros foram restringidos com base nos dados disponíveis ou ajustados a vários conjuntos de dados de séries temporais (consulte a Tabela Complementar 1). No total, 43 parâmetros foram restringidos e 61 parâmetros foram ajustados. Os parâmetros restritos foram estimados a partir de séries temporais dados dos mRNAs ou proteínas do relógio: LHY/CCA1 (g3, p1, p2, p3, m1, m2, m3, m4, c), TOC1 (g4, g5, m5, e), PRR9 (n4, m12, g8, g9, q3), PRR7 (m14), PRR5 (m16), ELF4 (g6), ELF3 (g16), LUX (m34), GI (g15, q2), ZTL (p13, m20, m21), COP1d e COP1n (n5, n6, n14, m27, m31, m33, p6, p15) e proteínas HY5 e HFR1 (p22, p30, m25, m28; m29; usado apenas para o parâmetro COP1 (informações suplementares de [Pokhilko et al., 2010]).

2.2.6 Motifs

As GRN controlam a expressão de genes. As redes de transcrição de microrganismos bem estudados parecem ser constituídas por um pequeno conjunto de padrões de regulação recorrentes, chamados “motifs” [Alon, 2007].

Recentemente, os mesmos “motifs” foram encontrados em diversos organismos, de bactérias a humanos, sugerindo que eles servem como blocos básicos de construção de redes de transcrição. “Motifs” em outras redes biológicas também são mencionados, incluindo sinalização e redes neurais [Alon, 2007].

A principal ideia é que cada “motif” possa executar tarefas específicas. Essas funções têm sido analisadas usando modelos matemáticos e testadas com experimentos dinâmicos em células vivas. Muitos experimentos com “motifs” têm sido realizados. Esses experimentos poderiam trazer maiores explicações sobre a dinâmica dos muitos sistemas em que cada “motif” aparece [Alon, 2007].

Um dos tipos de “motifs” descritos em [Alon, 2007] é o “feedforward loop (FFL)” e o utilizado neste trabalho mais especificamente é denominado “the coherent type-1 FFL (C1-FFL)”. Ele aparece em centenas de sistemas genéticos em *E. coli* e levedura, bem como em outros organismos. Este “motif” consiste em três genes: um regulador, X, que regula Y e o gene Z, que é regulado tanto por

X e Y. Porque cada uma das três interações regulatórias no “C1-FFL” pode ser ativação ou repressão [Alon, 2007].

Para entender a função do “C1-FFL”, é preciso entender como X e Y são integrados para regular o Promotor. Duas ‘funções de entrada’ comuns são um “AND gate”, em que X e Y são necessários para ativar Z, e uma “porta OR”, na qual a ligação de qualquer regulador é suficiente.

No entanto, muito do comportamento essencial de “C1-FFL” podem ser entendidos focalizando o estereótipo Portas “AND” e “OR” como mostrado na Figura 2.8 [Alon, 2007].

A Figura 2.8 representa o funcionamento da interação entre os TFs utilizando portas “AND” e “OR”. Nesta Figura, especificamente, tanto em 2.8a quanto em 2.8b, estão representados fatores externos que podem interferir na interação entre os TFs.

Neste trabalho utiliza-se o conceito de portas “AND” e “OR” de um circuito, que encontra-se na definição do “motif” “C1-FFL”, para desenvolver o código do simulador.

2.3 Sistemas Multiagente

Um agente é uma entidade de software que exibe um comportamento autônomo e pró-ativo orientado a objetivos, que está situado em algum ambiente sobre o qual é capaz de realizar ações para alcançar seus próprios objetivos de projeto e a partir do qual percebe alterações [Jennings, 1996, Wooldridge and Jennings, 1995].

Dá-se o nome de Sistema Multiagente (MAS - MultiAgent System) a um agrupamento de agentes que trabalham cooperativamente, cada um deles resolvendo parte do problema. Segundo [Jennings, 1996], MAS também refere-se à sub-área da Inteligência Artificial Distribuída (IAD), que investiga o comportamento de um conjunto de agentes autônomos objetivando a solução de um problema que está além das capacidades de um único agente [Jennings, 1996].

Em [Wooldridge and Jennings, 1995] considera-se um agente como sendo uma entidade com capacidade de resolução de problemas encapsulada. Inserido nesta visão, define o agente como tendo as seguintes propriedades:

- autonomia - executam a maior parte de suas ações sem interferência direta de agentes de outra natureza, possuindo controle total sobre suas ações e estado interno;
- habilidade social - por impossibilidade de resolução de certos problemas ou por outro tipo de conveniência, interagem com outros agentes, para completarem a resolução de seus problemas. Disto surge a necessidade de que os agentes tenham capacidade para comunicar seus requisitos aos outros e um mecanismo decisório interno que defina quando e quais interações são apropriadas;
- capacidade de reação - percebem e reagem às alterações no ambiente em que estiverem inseridos.
- capacidade pró-ativa - agentes do tipo deliberativo, além de atuar em resposta às alterações ocorridas em seu ambiente, apresentam um comportamento orientado a objetivos, tomando

iniciativas.

Tradicionalmente técnicas de simulação são baseadas em modelos matemáticos ou estocásticos, geralmente equações diferenciais que relacionam vários parâmetros e descrevem a dinâmica dos sistemas [Jennings, 1996].

Exemplificando, simulações ecológicas podem relacionar o tamanho da população de espécies específicas com o crescimento de diferentes espécies e o número de predações. Porém, há uma desvantagem nesses métodos, parâmetros de entrada e saída são definidos no mesmo nível, ou seja, eles não relacionam os parâmetros globais aos locais [Jennings, 1996]. Sendo assim, a partir de equações diferenciais, não é possível relacionar um parâmetro global, como o tamanho da população, a parâmetros locais, como os processos de decisão do indivíduo [Jennings, 1996].

Sabe-se que os MAS têm poucos parâmetros a serem ajustados, mantendo sua capacidade de representar adequadamente um modelo biológico. Assim, para prever valores que são difíceis de analisar na natureza, utiliza-se neste trabalho esse tipo de sistema, pois tratam-se de ferramentas muito flexíveis que permitem a representação de cada indivíduo presente em um sistema, facilitando, a avaliação de novas hipóteses no modelo [Jennings, 1996].

2.3.1 NetLogo

O NetLogo é um ambiente de modelagem programável para simular fenômenos naturais e sociais. Foi de autoria de Uri Wilensky em 1999 e, está em desenvolvimento contínuo desde então no Centro de Aprendizagem Conectada e Modelagem Baseada em Computador [Tisue and Wilensky, 2004].

O NetLogo é particularmente adequado para modelar sistemas complexos em desenvolvimento ao longo do tempo. Os modeladores podem dar instruções para centenas ou milhares de “agentes”, todos operando de forma independente. Isso torna possível explorar a conexão entre o comportamento em nível micro dos indivíduos e os padrões em nível macro que emergem de sua interação [Tisue and Wilensky, 2004].

O NetLogo possui um suporte de documentação e de tutoriais para os modeladores. Ele também vem com a Biblioteca de Modelos, uma grande coleção de simulações pré-escritas que podem ser usadas e modificadas. Essas simulações abordam áreas de conteúdo nas ciências naturais e sociais, incluindo biologia e medicina, física e química, matemática e ciência da computação e economia e psicologia social.

Na Figura 2.2 pode-se visualizar a Biblioteca de Modelos já mencionada no capítulo anterior, a Figura 2.3 mostra na aba interface um desses modelos que simula a disseminação de um vírus, nesta Figura a simulação não foi iniciada ainda, já na Figura 2.4 mostra-se a simulação em andamento.

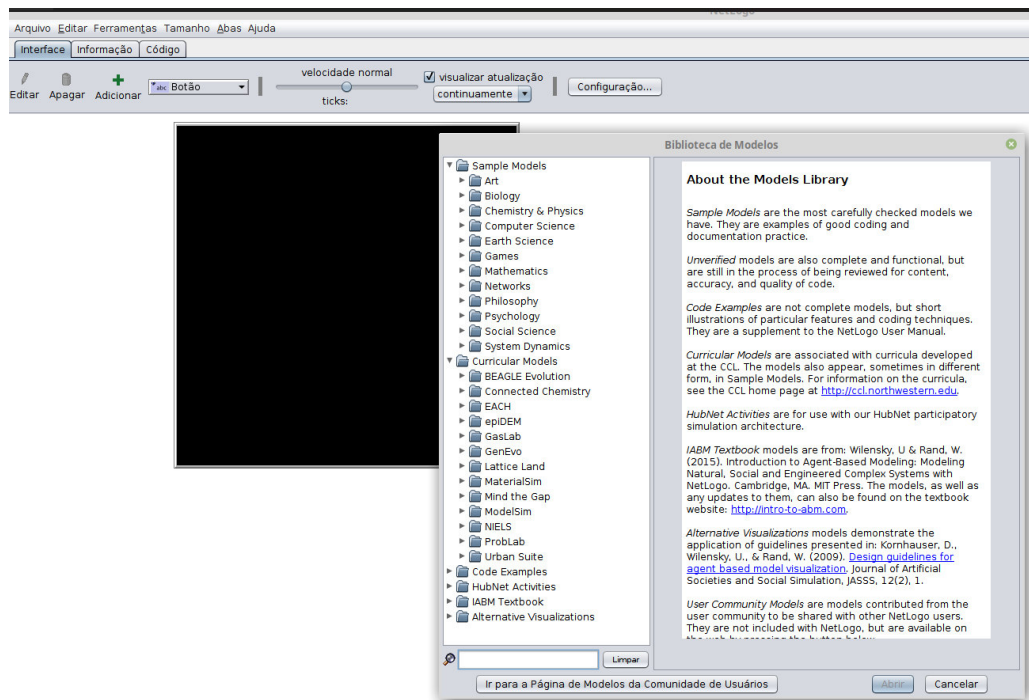


Figura 2.2: Biblioteca do Netlogo

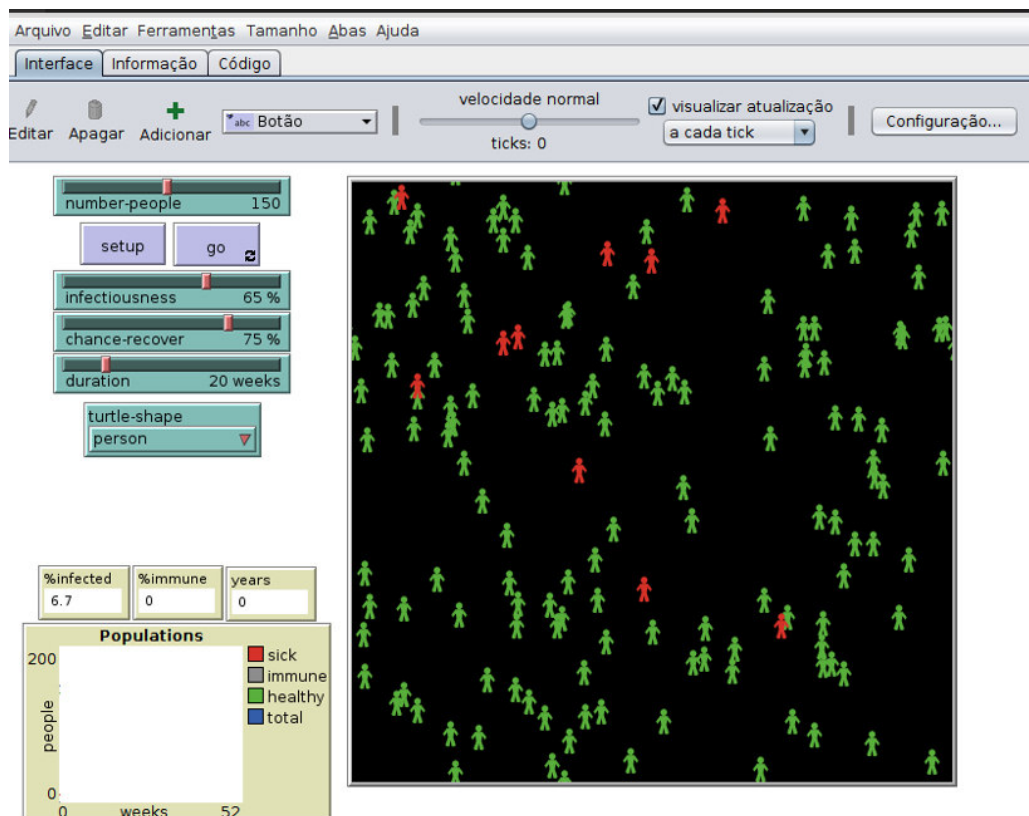


Figura 2.3: Modelo “Virus” antes de inicializar a simulação

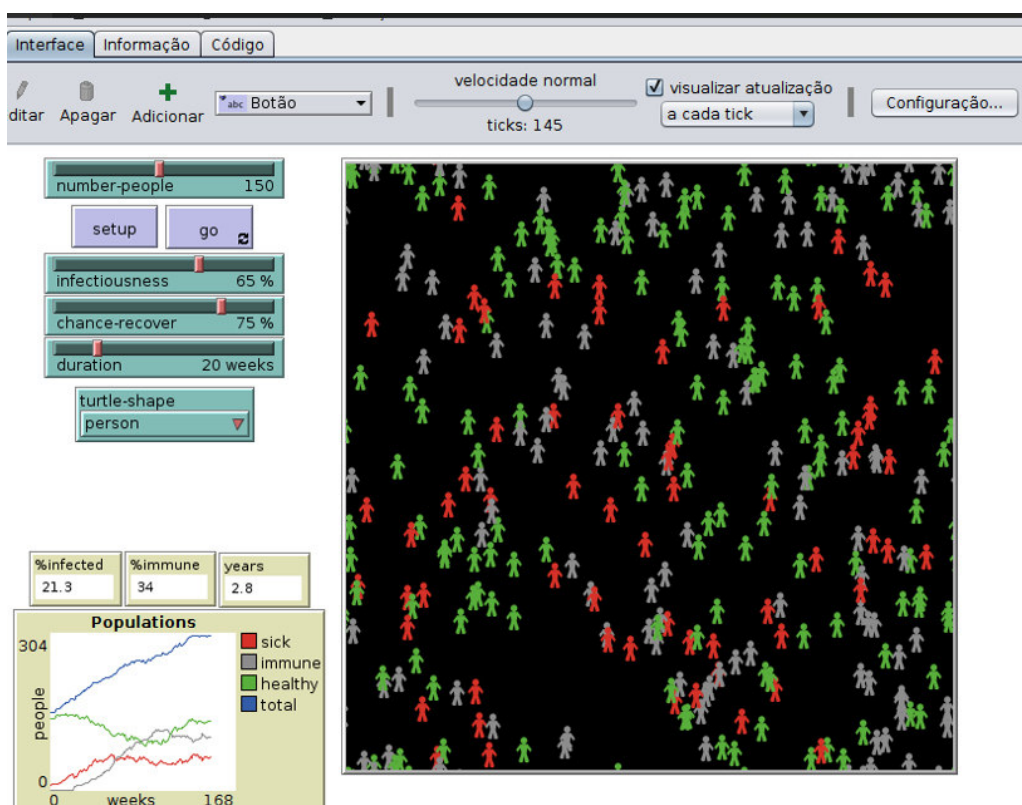
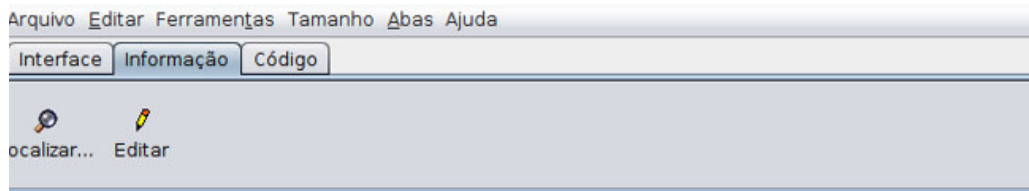


Figura 2.4: Modelo “Virus” com a simulação em andamento

Na Figura 2.5 verifica-se a existência da aba informação, que possui informações sobre o modelo desenvolvido, enquanto que na Figura 2.6 encontra-se a aba código que possui código do modelo em destaque.



WHAT IS IT?

This model simulates the transmission and perpetuation of a virus in a human population.

Ecological biologists have suggested a number of factors which may influence the survival of a directly transmitted virus within a population. (Yorke, et al. "Seasonality and the requirements for perpetuation and eradication of viruses in populations." Journal of Epidemiology, volume 109, pages 103-123)

HOW IT WORKS

The model is initialized with 150 people, of which 10 are infected. People move randomly about the world in one of three states: healthy but susceptible to infection (green), sick and infectious (red), and healthy and immune (gray). People may die of infection or old age. When the population dips below the environment's "carrying capacity" (set at 300 in this model) healthy people may produce healthy (but susceptible) offspring.

Some of these factors are summarized below with an explanation of how each one is treated in this model.

The density of the population

Population density affects how often infected, immune and susceptible individuals come into contact with each other. You can change the size of the initial population through the NUMBER-PEOPLE slider.

Population turnover

As individuals die, some who die will be infected, some will be susceptible and some will be immune. All the new individuals who are born, replacing those who die, will be susceptible. People may die from the virus, the chances of which are determined by the slider CHANCE-RECOVER, or they may die of old age.

In this model, people die of old age at the age of 50 years. Reproduction rate is constant in this model. Each turn, if the carrying capacity hasn't been reached, every healthy individual has a 1% chance to reproduce.

Degree of immunity

If a person has been infected and recovered, how immune are they to the virus? We often assume that immunity lasts a lifetime and is assured, but in some cases immunity wears off in time and immunity might not be absolutely secure. In this model

Figura 2.5: Aba informação no Netlogo

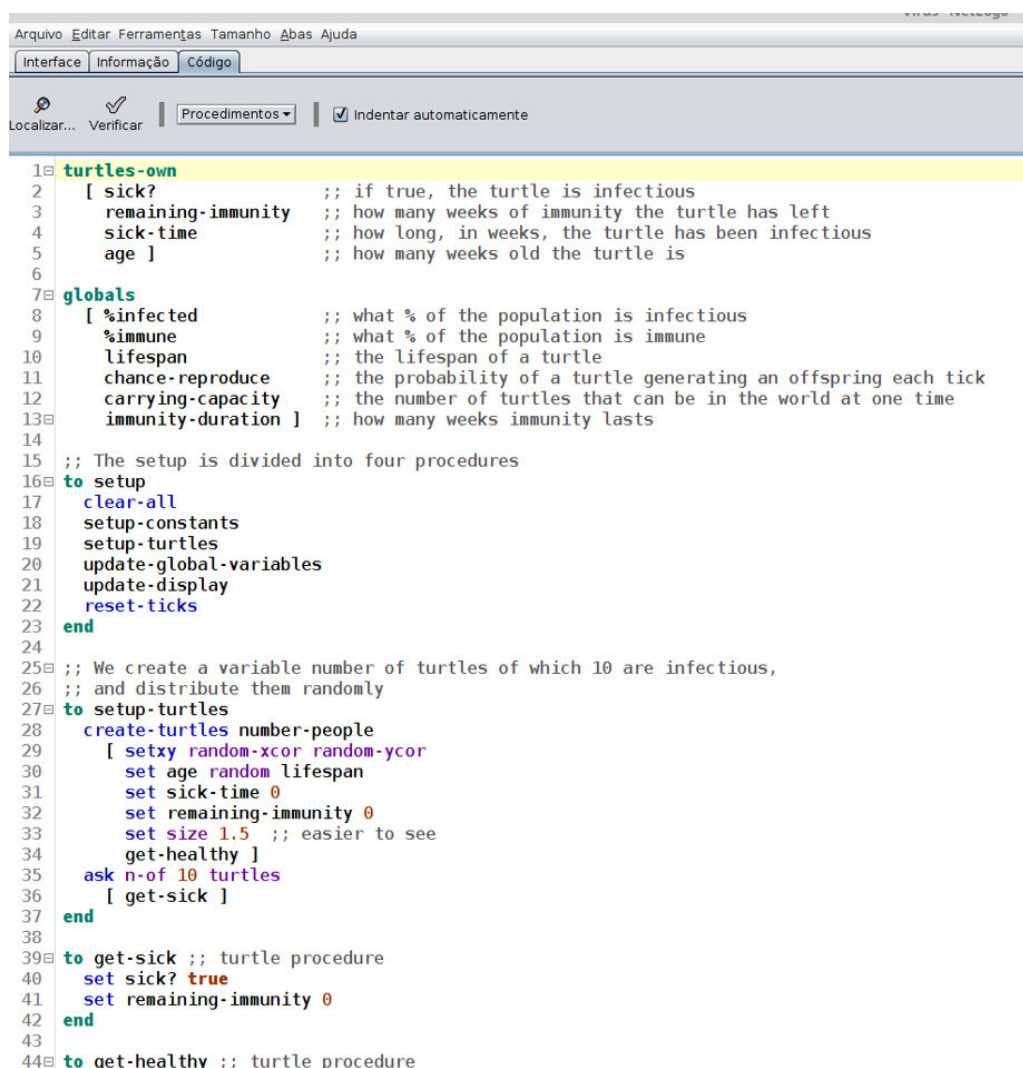


Figura 2.6: Aba código no Netlogo

O NetLogo é a próxima geração da série de linguagens de modelagem multiagente, incluindo StarLogo e StarLogoT. O NetLogo é executado na Java Virtual Machine (JVM) que é uma máquina de computação abstrata ou interface de máquina virtual que conduz o código java², portanto, funciona em todas as principais plataformas (Mac, Windows, Linux, etc.). É executado como um aplicativo de desktop [Tisue and Wilensky, 2004].

No contexto da Bioinformática, pode-se verificar que o NetLogo tem sido usado extensivamente para promover nossa compreensão dos sistemas em várias disciplinas diferentes [Chiacchio et al., 2014].

Por exemplo, várias aplicações de Netlogo para modelagem da dinâmica do sistema imunológico, assim como o trabalho realizado que desenvolveu um modelo para estudar a resposta de células inflamatórias (macrófagos) e células envolvidas na remodelação (fibroblastos) à exposição a partículas [Brown et al., 2011]. Esse modelo NetLogo, mesmo com limitações claras (estruturas teciduais variadas no pulmão, ausência de complemento total de células e outras células imunes importantes que desempenham um papel no processo inflamatório), foi capaz de prever a existência de aspectos bi-

²<https://www.w3schools.in/java-tutorial/java-virtual-machine/>

ologicamente relevantes, ou seja, cicatrização e retorno à linha de base, dano ao tecido localizado e fibrose e dano extenso e fibrose [Brown et al., 2011].

Um modelo Netlogo, que descreve o cenário típico observado na esclerose múltipla remitente-recorrente (o tipo mais relevante de EM), é apresentado em [Pennisi et al., 2013]. O modelo leva em consideração duas populações que representam linfócitos CD8 específicos que são a causa de dano neural e linfócitos T reguladores específicos, que podem falhar em suprimir a expansão de CD8 [Pennisi et al., 2013]. O modelo mostrou capacidade de capturar a dinâmica essencial da EM remitente-recorrente, simulando a ausência e a presença de disfunções dos mecanismos de equilíbrio cruzado CD8-CD4-Treg em nível local. Em particular, sugeriu que a presença de uma predisposição genética nem sempre é uma condição suficiente para o desenvolvimento da doença [Pennisi et al., 2013].

Outro exemplo que pode ser descrito, apresenta um modelo de simulação de computador que usa um sistema baseado em agente desenvolvido com NetLogo para imitar o desenvolvimento e progressão de tumores sólidos [Dreau et al., 2009]. O modelo inclui as próprias características do tumor, a resposta imune do hospedeiro e o nível de vascularização do tumor. Simulações realizadas indicam a importância fundamental das necessidades de nutrientes das células tumorais e da capacidade de resposta inicial do sistema imunológico na progressão tumoral [Dreau et al., 2009].

2.4 Base teórica para os Estudos de Caso

2.4.1 Processo de Michaelis-Menten

O processo de regulação entre os genes emerge das interações nos Sítios de Ligação e podem ser modelado como um processo de cinética enzimática, também conhecido como cinética de Michaelis-Menten [Johnson and Goody, 2011].

O processo de Michaelis-Menten foi definido há mais de 100 anos e mostra que a taxa de reação catalizada para uma enzima é proporcional à concentração do complexo substrato-enzima, conforme Equação 1.

$$v = \frac{d[P]}{dt} = \frac{V_{max}[S]}{K_M + [S]} \quad (1)$$

A taxa de reação varia linearmente com a concentração de substrato [S] (cinética de primeira ordem). No entanto, em níveis mais altos [S] com $[S] \gg K_M$, a reação se torna independente de [S] (cinética de ordem zero) e se aproxima assintoticamente de sua taxa máxima V_{max} [Johnson and Goody, 2011].

Michaelis e Menten mediram a velocidade inicial em função da concentração de sacarose e assim derivaram a Equação 1. A Figura 2.7a mostra o comportamento da equação de Michaelis-Menten em termos de taxa de reação e a Figura 2.7b mostra as concentrações do substrato e do produto para o mesmo processo. Essas cinéticas auxiliam como primeira referência para nossos estudos de simulação.

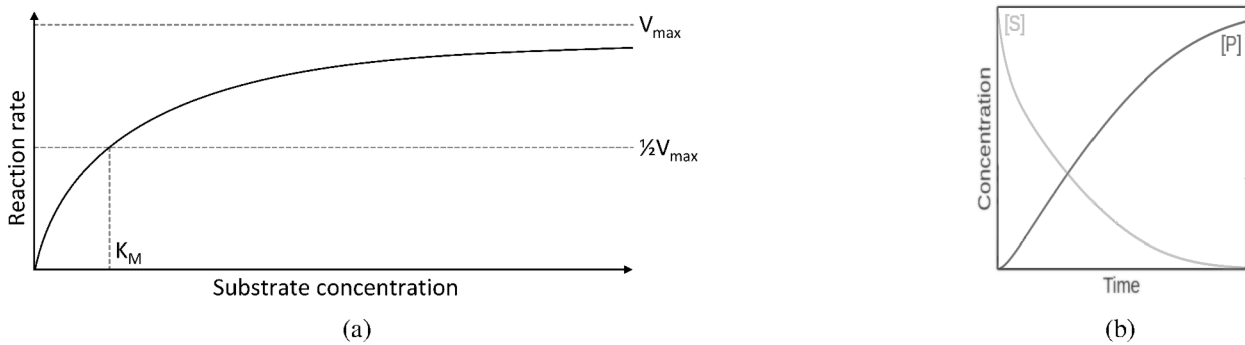


Figura 2.7: (a) curva típica da taxa de reação Michaelis-Menten. (b) variação típica na concentração de substrato seguindo uma abordagem de Michaelis-Menten [Johnson and Goody, 2011].

2.4.2 Uri Alon

As redes de regulatórias de transcrição controlam a expressão de genes. As redes de transcrição de microrganismos em estudo parecem ser constituídas por um pequeno conjunto de padrões de regulação recorrentes, chamados *motifs* de rede. Os mesmos *motifs* de rede foram encontrados recentemente em diversos organismos, de bactérias a humanos, sugerindo que eles servem como blocos básicos de construção de redes de transcrição. Em [Alon, 2007], são apresentados os *motifs* da rede e suas funções, com ênfase em estudos experimentais.

Um dos “motifs” mais estudado neste artigo é o C1-FFL que é um elemento de “atraso sensível ao sinal” e um detector de persistência. No C1-FFL, ambos X e Y são ativadores de transcrição [Alon, 2007].

Com uma função de entrada AND, o C1-FFL mostra um atraso após a estimulação, mas nenhum atraso na estimulação é realizado. Para ver isso, é preciso seguir o comportamento do FFL. Quando o sinal S_x aparece, X fica ativo e liga rapidamente seus promotores. Como um resultado, Y começa a acumular-se. No entanto, devido à Função de entrada AND, a produção Z começa apenas quando concentração de Y ultrapassa o limiar de ativação para o Promotor Z. Isso resulta em um atraso na expressão Z considerando S_x [Alon, 2007].



Figura 2.8: (a) “motif C1-FFL” com uma função “AND”. (b) “motif C1-FFL” com uma função “OR” [Alon, 2007].

Como representado nas Figuras 2.8 (a) e (b), os “motifs” podem ser considerados circuitos recorrentes de interações a partir dos quais as redes são construídas. Os “motifs” foram definidos sistematicamente pela primeira vez em *Escherichia coli*, na qual foram detectados como padrões que ocorreram na rede de transcrição com muito mais frequência do que seria esperado em redes aleatórias. Os mesmos “motifs” foram encontrados desde então em organismos, desde bactérias e leveduras a plantas e animais [Alon, 2007].

Sendo assim, o trabalho de Uri Alon foi utilizado como base para o desenvolvimento desta ferramenta e também utilizado como Estudo de Caso para validação do desenvolvimento da mesma.

2.4.3 Ciclo circadiano Eucariótico da Planta *Arabidopsis Thaliana*

Esta seção apresenta o estudo de caso a ser utilizado, da modelagem do ciclo circadiano eucariótico da planta *Arabidopsis thaliana*, que foi apresentado em “Data assimilation constrains new connections and components in a complex, eukaryotic circadian clock model” [Pokhilko et al., 2010].

Os ritmos circadianos são difundidos na biologia de quase todos os organismos eucarióticos, incluindo plantas. As oscilações auto-sustentáveis de 24h dos genes do relógio central conduzem a uma expressão rítmica do genoma de *Arabidopsis*, fornecendo assim, uma vantagem adaptativa para antecipar as mudanças diárias nas condições ambientais. O conceito atual do ciclo circadiano de *Arabidopsis* é de uma rede genética que consiste em loops de feedback matinais e noturnos interconectados por um loop central [Pokhilko et al., 2010].

Neste trabalho, os genes expressos pela madrugada LATE ELONGATED HYPOCOTYL (LHY) e CIRCADIAN CLOCK ASSOCIATED 1 (CCA1) formam um loop de feedback negativo matinal através da ativação de seus inibidores, PSEUDO-RESPONSE REGULATOR 9 (PRR9) e PRR7. LHY e CCA1 também controlam a fase dos genes noturnos, inibindo sua expressão pela manhã. O importante gene noturno TIMING OF CAB EXPRESSION 1 (TOC1), que é expresso ao entardecer, regula-se no loop noturno do relógio através da inibição de seu hipotético ativador, o gene Y . O gene

expresso noturno GIGANTEA (GI) foi sugerido anteriormente ser um componente de Y. O loop da noite realimenta o loop da manhã através do hipotético gene X, que ativa a expressão de LHY/CCA1 no final da noite. A luz estimula a expressão de LHY/CCA1, PRR9 e Y [Pokhilko et al., 2010].

A proteína GI regula o ciclo pela estabilização da proteína F-box ZEITLUPE (ZTL) na presença de luz. A ZTL, por sua vez, direciona a proteína TOC1 à degradação pelo proteossoma. Os resultados de mutantes gi questionaram a implicação de que GI regula diretamente TOC1. O gene expresso à noite PRR5 não foi explicitamente incluído no modelo, embora juntamente com os genes da manhã PRR9 e PRR7, PRR5 seja importante para a regulação da expressão de LHY e CCA1 [Pokhilko et al., 2010].

Em contraste com o modelos anteriores, o grupo de pesquisadores de [Pokhilko et al., 2010] identificou dados que restringiam 35 dos 90 parâmetros do modelo. Os restantes parâmetros foram ajustados a dois tipos de dados: os perfis quantitativos dos componentes moleculares do ciclo e os valores dos períodos de “freerunning” dos ritmos circadianos, em plantas de tipo selvagem (wt) e LHY/CCA1, TOC1, ZTL e PRR7/PRR9 mutantes, sob condições ambientais variadas. O grupo aplicou o modelo para analisar os mecanismos de regulação circadiana sob perturbações genéticas e ambientais adicionais.

O modelo proposto deste trabalho é apresentado na Figura 2.9 e o modelo detalhado encontra-se na Figura 2.10. Trata-se de um modelo ODE (Ordinary Differential Equation - Equação Diferencial Ordinária), que consiste em 28 equações diferenciais ordinárias e 104 parâmetros. Os valores de 43 parâmetros foram fixados com base nos dados disponíveis e 61 parâmetros foram ajustados a vários conjuntos de dados de séries temporais.

Uma descrição detalhada do modelo ODE é apresentada nas informações complementares do trabalho [Pokhilko et al., 2010].

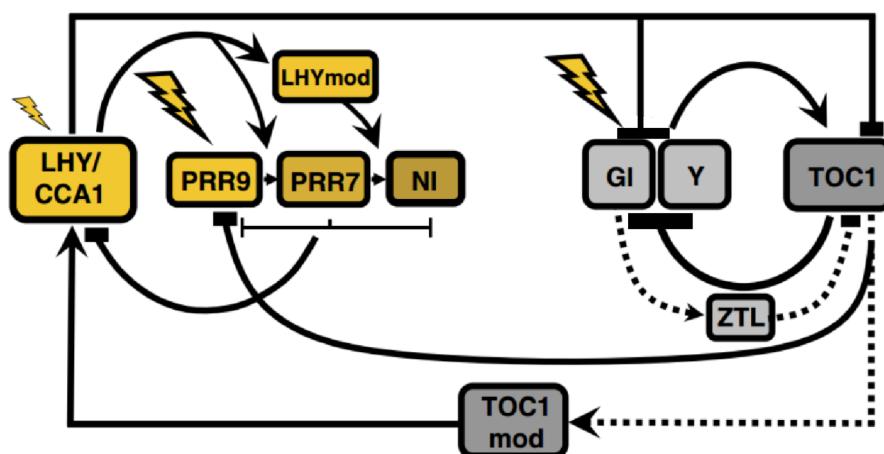


Figura 2.9: Modelo da *Arabidopsis thaliana*. Os principais elementos do modelo de relógio circadiano de *Arabidopsis* estendido. Neste modelo, Elementos da manhã e osciladores noturnos são mostrados em amarelo e cinza, respectivamente. As proteínas são mostradas apenas para ZTL, LHY modificado (LHYmod) e TOC1 modificado (TOC1mod). Interações genéticas, setas sólidas; regulação pós-traducional, setas tracejadas. As entradas de luz para a transcrição do gene são marcadas por “flashes” [Pokhilko et al., 2010]. O esquema completo do modelo é apresentado na Figura 2.10.

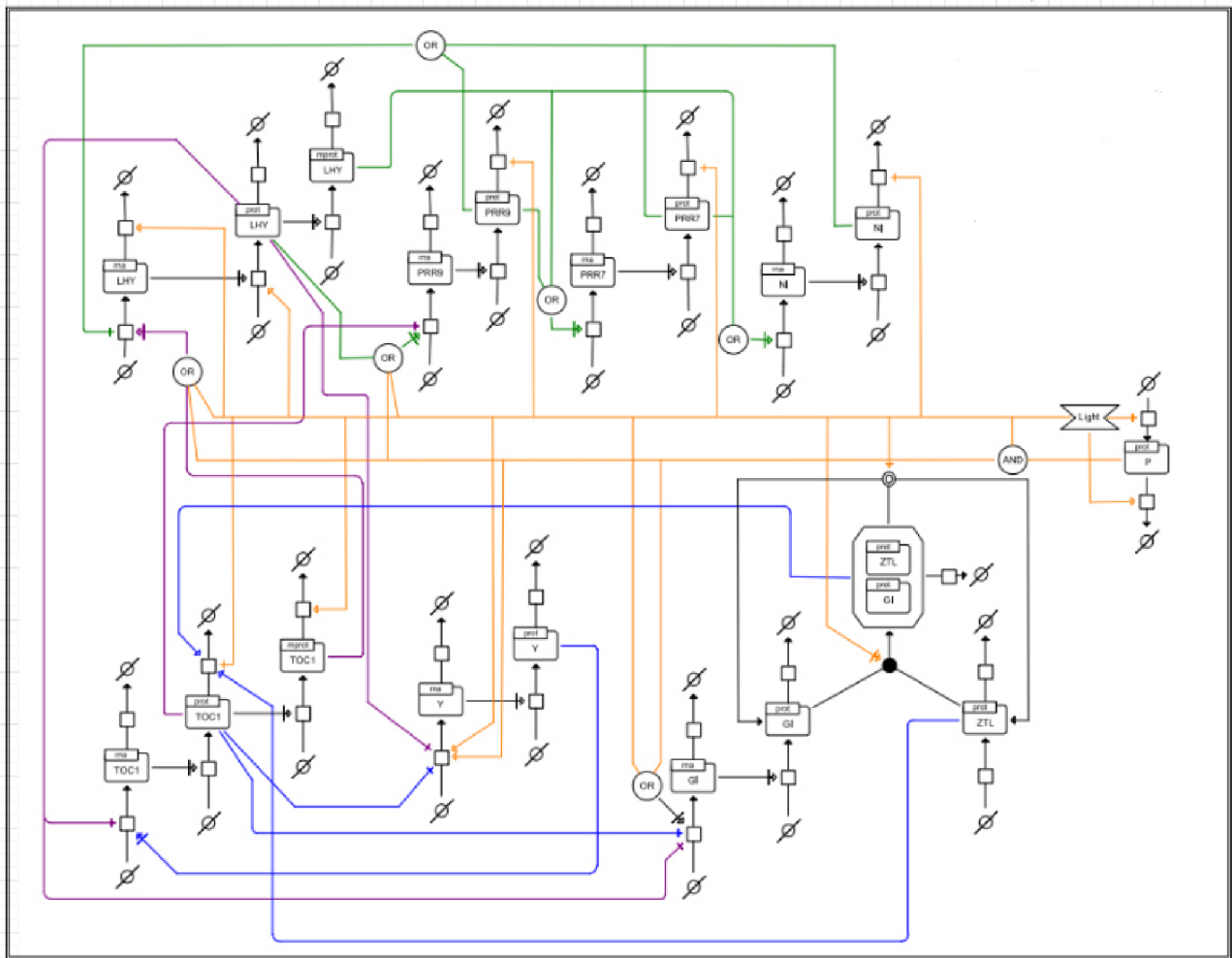


Figura 2.10: As linhas verde, azul e violeta representam os ciclos matutinos e vespertinos e as conexões entre eles. As linhas laranja mostram as entradas de luz [Pokhilko et al., 2010].

O modelo descreveu a estabilização da proteína ZTL em complexo com a proteína GI e a aceleração da degradação da proteína TOC1 por ZTL. Assim, o GI inibe a função de TOC1 no modelo ao nível da proteína TOC1, através da regulação positiva de ZTL. Juntamente com a inibição de Y [Pokhilko et al., 2010].

A expressão rítmica estendida de TOC1 pela luz no gene LHY/CCA1, juntamente com a ausência de ativação direta da expressão de TOC1 pela luz, implicou que Y era o componente do ciclo noturno. Além disso, a expressão noturna de TOC1 em plantas wt sugeriu que a expressão de Y foi retardada através da inibição pela proteína LHY/CCA1. O primeiro pico de Y após o amanhecer é causado pela ativação aguda da expressão Y pela luz, enquanto um segundo pico maior 10 h após o amanhecer está relacionado à liberação da expressão Y a partir de sua inibição pela proteína LHY/CCA1. A ampla expressão de Y durante o período de luz permite uma medida de “sensação de crepúsculo” pelo loop noturno do ciclo [Pokhilko et al., 2010].

A proteína GI regula o ciclo no nível pós-tradução estabilizando a proteína F-box ZEITLUPE (ZTL) na presença de luz. A ZTL, por sua vez, é necessária para o direcionamento da proteína TOC1 à degradação pelo proteossoma [Pokhilko et al., 2010].

3 REVISÃO SISTEMÁTICA

As revisões que pode-se considerar como sistemáticas começaram a aparecer na década de 1950. No entanto, o desenvolvimento metodológico dessas pesquisas na área da saúde, por exemplo, só se consolidou ao fim da década de 1980 [Galvao and Pereira, 2014].

Foram marcos nesse processo a publicação do livro “Effective Care During Pregnancy And Child-birth” e, na década seguinte, a criação da Colaboração Cochrane, em Oxford, Reino Unido. As revisões sistemáticas da literatura (SLR - Systematic Literature Review) devem ser abrangentes e não tendenciosas na sua preparação [Galvao and Pereira, 2014].

Os critérios adotados são divulgados de modo que outros pesquisadores possam repetir o procedimento. SLRs de boa qualidade são consideradas o melhor nível de evidência para tomadas de decisão [Galvao and Pereira, 2014].

Uma boa maneira de esclarecer controvérsias é os melhores estudos de qualidade na área em questão. Dessa forma, a SLR é um tipo de pesquisa focada em um assunto bem definido, que visa identificar, selecionar, avaliar e sintetizar evidências de um tópico de pesquisa específico [Galvao and Pereira, 2014].

Em [Mariano et al., 2017], os autores propuseram uma padronização de etapas e procedimentos para o desenvolvimento desse tipo de pesquisa. O objetivo da SLR é criar uma visão geral de um problema específico e fornecer um resumo da literatura, garantindo padronização e precisão. Esse procedimento se torna um passo importante para o desenvolvimento do trabalho, pois pode ajudar a obter uma visão imparcial dos estudos publicados em diversas fontes (conferências e periódicos), além de robustez e segurança.

As etapas para executar uma SLR após uma padronização estão descritas em [Mariano et al., 2017] e são apresentadas na Figura 3.3. Devido ao fato de não existir uma norma que padronize o processo de SLRs na área de bioinformática, foi criado um padrão para o desenvolvimento destas normas que encontram-se descrito em [Mariano et al., 2017].

Numa etapa de preparação é realizada a definição da questão principal, a qual ajudará a definir as palavras-chave usadas para pesquisar nos bancos de dados. É aconselhado o uso de estratégias de busca, como busca por expressões completas, pois bancos de dados diferentes podem apresentar vários métodos para pesquisa. Após essa fase inicial o padrão para o desenvolvimento da SLR segue quatro sub-passos: avaliação do título, avaliação do resumo, leitura diagonal e leitura do texto completo.

Neste processo da SLR, é comum ser encontrada uma grande quantidade de artigos envolvendo as palavras-chave. Devido a esse fato, também é frequente o retorno para etapas anteriores para redefinir palavras-chave ou as suas variantes e o modo como são inseridas nas buscas, já que o processo é um ciclo e portanto é necessário “recalibrá-lo” em alguns casos de acordo com a avaliação prévia dos resultados.

Os filtros da etapa de leitura de resumos “alimentarão” a próxima etapa que é a de leitura diagonal. Nessa etapa e na próxima, que é de leitura completa do artigo, será necessária a atuação de um

pesquisador da outra área envolvida (multiagente).

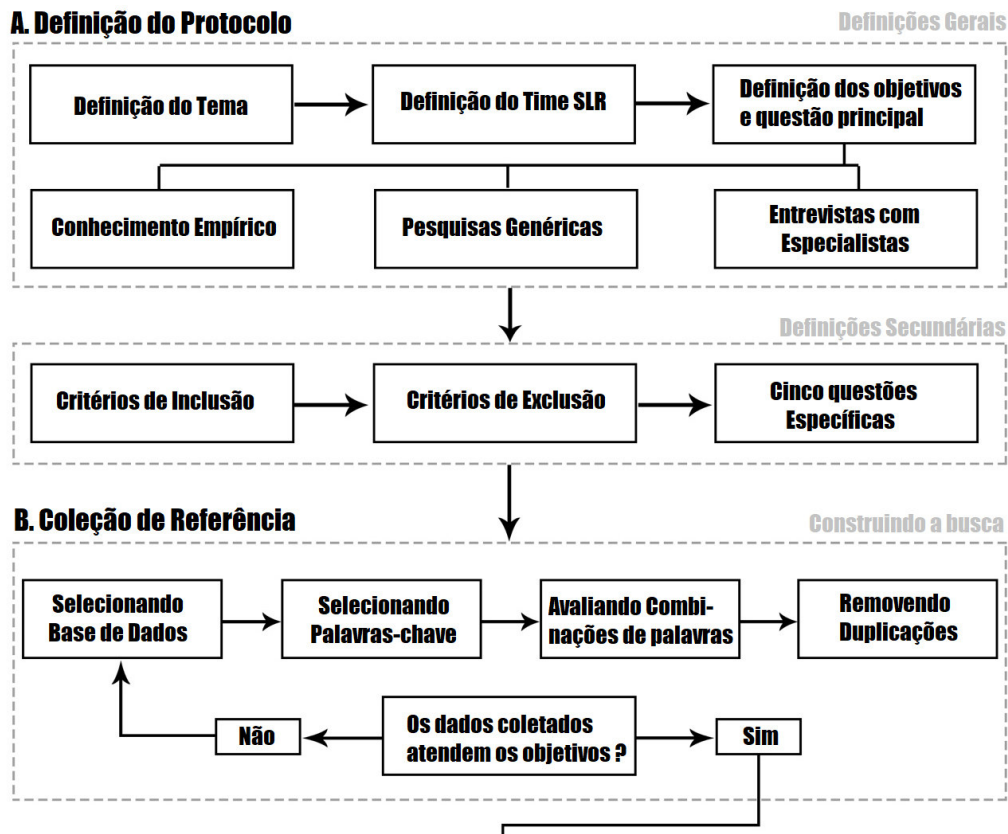


Figura 3.1: SLR Detalhado Adaptado de [Mariano et al., 2017] - Na etapa A é definido o protocolo a ser seguido. Na etapa B são definidas as bases de dados, as palavras-chave, realizada a avaliação das combinações das palavras chaves e por último são removidas as duplicações. Se o resultado for satisfatório segue-se para etapa C, senão segue-se para uma reavaliação do protocolo.

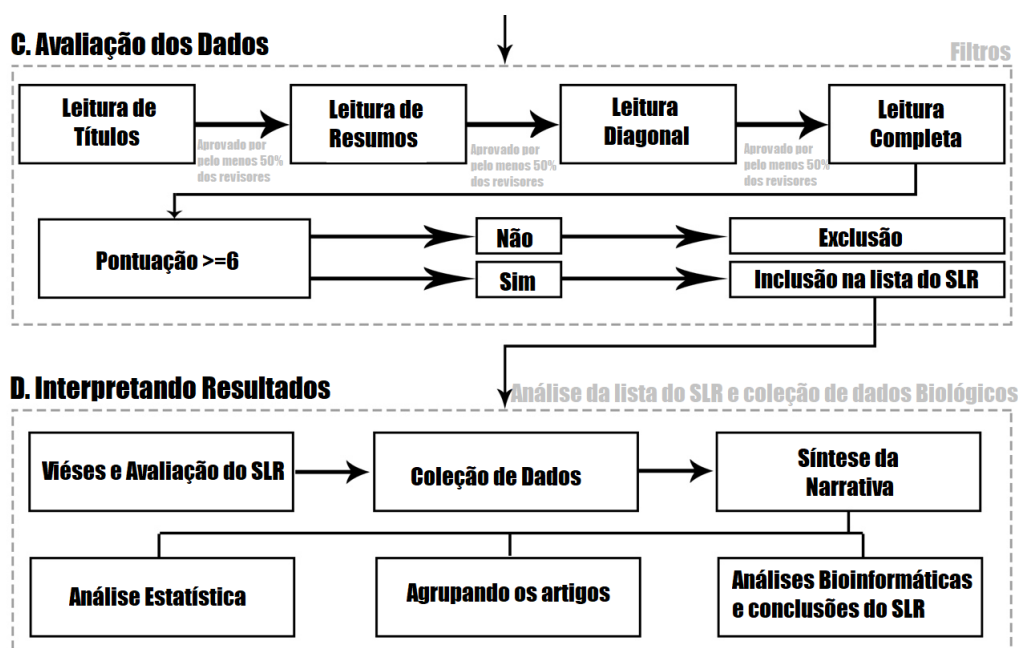


Figura 3.2: SLR Detalhado - Adaptado de [Mariano et al., 2017] - Na etapa C são realizados os filtros por título, resumo, leitura diagonal e leitura completa. Já na etapa D são realizadas as avaliações dos resultados obtidos.

Com esta SLR, objetiva-se obter como embasamento para o desenvolvimento de um ambiente multiagente para redes regulatórias, artigos resultantes da compilação de conteúdos científicos relevantes para as áreas de redes regulatórias e Sistemas Multiagente. E assim também garantir o ineditismo do trabalho.

Numa etapa de preparação é realizada a definição da questão principal: **Existe sistema multiagente para redes regulatórias?** E essa questão principal ajudará a definir as palavras-chave usadas para pesquisar nos bancos de dados.

Neste contexto, já foi delineada a questão principal, definidas as palavras-chave e realizadas as pesquisas de artigos nas seguintes bases de dados: Pubmed³, PMC⁴, Science Direct⁵, Scopus⁶ e Scielo⁷ e as palavras-chave definidas foram: “regulatory network” AND “multiagent system”. Porém, estas palavras-chave podem ter variantes na escrita e as buscas nas bases de dados são realizadas de maneiras diferentes, podendo variar também os operadores relacionais e expressões regulares [Agostinho et al., 2020a].

Num primeiro momento, foram utilizadas palavras chaves como “network and multiagents”, o que gerou como resultado o retorno de uma grande quantidade de artigos contendo essas palavras em seus títulos ou definição de palavras chaves do artigo. Por exemplo, nessa primeira realização foi incluída a base de dados Google Acadêmico e com essas palavras chaves chegaram a ser encontrados 19.700 artigos, indicando que seria necessário redefinir as palavras-chave e provavelmente também

³<https://www.ncbi.nlm.nih.gov/pubmed/>

⁴<https://www.ncbi.nlm.nih.gov/pmc/>

⁵<https://www.sciencedirect.com/>

⁶<https://www.scopus.com/home.uri>

⁷<https://www.scielo.org/>

as bases de dados.

O Protocolo de definições apresenta questão principal, objetivo, critérios de inclusão, critério de exclusão e questões específicas. O protocolo completo no qual foi baseado o protocolo para esta revisão está representado nas Figuras 3.1 e 3.2. O protocolo para este trabalho é apresentado na Tabela 1 e as palavras chaves utilizadas com suas variantes encontram-se na Tabela 2 [Agostinho et al., 2020a]. Os resultados de cada etapa da SLR realizada abrangem um período delimitador até o ano 2018.

Tabela 1: Protocolo de definição da SLR (Baseado em [Mariano et al., 2017])

Questão principal	Existe um Sistema multiagente para redes regulatórias?
Objetivo	O objetivo é encontrar pesquisas que envolvam Sistemas Multiagente e redes regulatórias
Critérios de Inclusão	Os critérios de inclusão são: • Estudos que mencionem Sistemas Multiagente e (redes regulatórias ou redes genéticas); • Estudos que contenham as palavras-chave: redes regulatórias, Multiagente, Multiagent Systems.
Critérios de Exclusão	Os critérios de exclusão são: • O estudo não é inglês; • Os artigos não contém nenhum estudo sobre (redes regulatórias ou redes genéticas) e (Multiagente or Multiagente Sistemas).
Questões específicas	Questões específicas: (i) É um artigo que contenha pesquisa sobre redes regulatórias e sistemas multigente? (ii) É um artigo que contenha pesquisa sobre redes regulatórias e Multiagente? (iii) É um artigo que contenha pesquisa sobre vias regulatórias e Multiagente ? (iv) É um artigo que contenha pesquisa sobre vias regulatórias e Sistemas Multiagente ?

Tabela 2: Palavras Chaves

Base	Chave
PMC	(regulatory network AND multiagents systems)
PMC	regulatory network AND multiagents
PMC	regulatory network AND multiagents system
PMC	regulatory network AND multiagent system
PMC	regulatory network AND multiagent systems
PMC semaspas	(regulatory pathways AND multiagent)
PMC semaspas	(regulatory pathways AND multiagents)
PMC semaspas	(regulatory Pathways AND multiagents system)
PMC semaspas	(regulatory pathways AND Multiagent System)
PMC semaspas	(regulatory pathways AND multiagent systems)
PMC	regulatory network AND multiagent
Pubmed	(regulatory network AND multiagent)
Pubmed	(regulatory network AND multiagents system)
Pubmed	(regulatory network AND multiagent systems)
Pubmed	(regulatory pathways AND multiagent system)
science	(regulatory network) AND multiagent
science	(regulatory network) AND multiagent
science	(regulatory network) AND multiagents System
science	(regulatory network) AND multiagents Systems
science	(regulatory network) AND multiagent System
science	(regulatory network) AND multiagent Systems
scopus	regulatory network AND "multiagent system"
scopus semaspas	(regulatory pathways AND multiagent)

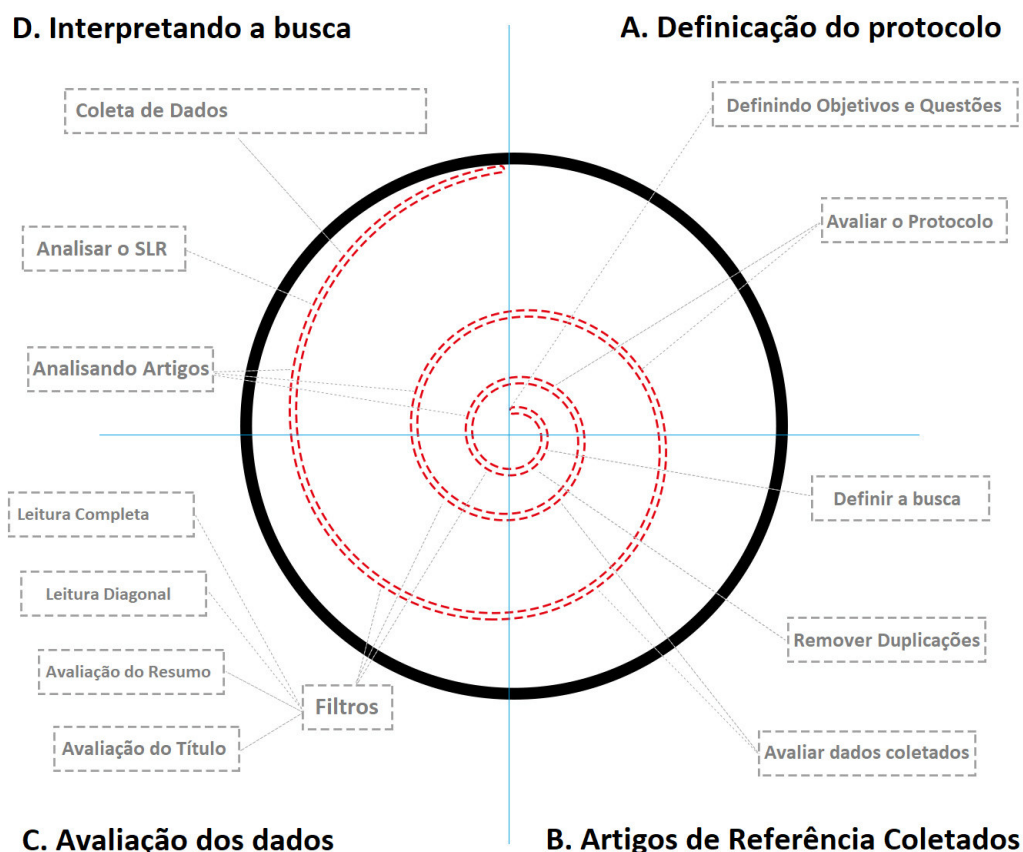


Figura 3.3: Modelo em espiral da revisão [Mariano et al., 2017]

Depois de definidas essas questões, então é feita a pesquisa para obtenção da literatura, verificadas as duplicações e feita a preparação para as etapas de avaliação. Para a obtenção dos artigos e materiais foram utilizadas as ferramentas Mendeley⁸ e JabRef⁹.

Na etapa de filtros, primeiramente filtra-se por título e a segunda etapa é a filtragem por resumos. Até a etapa de filtro por resumos é necessário apenas um pesquisador, nas demais o material coletado é avaliado por três pesquisadores, dois da área de bioinformática e um da área de multiagente [Agostinho et al., 2020a].

A etapa seguinte é de leitura diagonal, onde cada um dos pesquisadores devem ler a introdução, os títulos das Figuras e tabelas e as conclusões dos trabalhos classificados para essa etapa. Na leitura diagonal, os revisores podem aprimorar seus conhecimentos sobre o tema estudado. Assim, nesta etapa, os revisores devem verificar se as questões específicas atendem ao objetivo estabelecido no protocolo inicial. Nesta etapa espera-se que reduza drasticamente a quantidade de trabalhos a serem avaliados na próxima etapa [Agostinho et al., 2020a].

Após a realização da leitura diagonal é realizada a etapa de leitura completa do texto que definirá quais trabalhos ficarão para serem incluídos como base do estado da arte deste tema. Em [Mariano et al., 2017], é esquematizado esse processo baseado no padrão SLR.

A Tabela 3 apresenta o resumo dos resultados por etapas de filtro. As colunas A, B e C representam

⁸<https://www.mendeley.com/>

⁹<https://www.jabref.org/>

a aprovação ou reprovação dos artigos avaliados por cada um dos pesquisadores. Os valores nas colunas A, B e C podem ser 0, 1 ou ?, sendo que 0 significa que o artigo foi reprovado para próxima etapa de filtro, 1 significa aprovado e ? indica que é preciso ser feita outra análise pelos pesquisadores envolvidos na avaliação da SLR.

Tabela 3: Resultados (resumido) após etapa de filtro por resumo

A	B	C	Título
0	1	1	Use of evolved artificial regulatory networks to simulate 3D cell differentiation
0	1	1	A cell pattern generation model based on an extended artificial regulatory network
1	?	0	Agent-Based Modeling of the Immune System: NetLogo, a Promising Framework
1	0	?	Combining biochemical network motifs within an ARN-agent control system
1	1	?	Genetic regulatory network inference using Recurrent Neural Networks trained by a multi agent system
1	1	1	BioASF: A framework for automatically generating executable pathway models specified in BioPAX
0	?	?	Lasso granger causal models: Some strategies and their efficiency for gene expression regulatory networks
1	1	1	Consensus of multiagent systems in the cooperation–competition network with inherent nonlinear dynamics: A time-delayed control approach
0	0	1	State feedback stabilization for probabilistic Boolean networks
1	1	?	A Dynamic Multiagent Genetic Algorithm for Gene Regulatory Network Reconstruction Based on Fuzzy Cognitive Maps
0	1	1	A time series driven decomposed evolutionary optimization approach for reconstructing large-scale gene regulatory networks based on fuzzy cognitive maps
0	0	0	Detection of driver metabolites in the human liver metabolic network using structural controllability analysis

Nas Tabelas 4 e 5 são descritos, resumidamente, os artigos resultantes após as etapas de filtro por título, resumo, leitura diagonal e leitura completa [Agostinho et al., 2020a].

O artigo [Liu et al., 2016] tem como objetivo reconstruir GRN a partir de perfis de expressão de séries temporais com base em mapas cognitivos fuzzy (Fuzzy Cognitive Maps - FCMs).

Em [Khan et al., 2003] descreve-se a simulação de redes de transdução de sinal (ST) usando a arquitetura DECAF MAS.

Em [Haydarlou et al., 2016] é introduzida uma estrutura de simulação genérica para o BioPAX , em [Ghazikhani et al., 2011] é proposto um novo algoritmo para inferência de rede regulatória genética.

Em [Sanfilippo et al., 2012] apresenta-se uma abordagem em que a modelagem de emergência seja estendida às redes regulatórias e demonstre sua aplicação na descoberta de vias neuroprotetoras.

Em [Pham, 2008] é proposta uma abordagem baseada em vários agentes para simular grandes redes booleanas aleatórias, e em [Yang and Sun, 2011] apresenta-se um novo método para reconstruir a rede usando Sistemas Multiagente para fundir os dados de expressão gênica e os dados de ligação ao Fator de Transcrição (TF).

Tabela 4: Resultado da SLR - Comparação entre artigos - Parte I

Artigo	Objetivo	Técnicas	Resultados
[Liu et al., 2016]	Reconstruir GRN a partir de perfis de expressão de séries temporais com base em mapas cognitivos fuzzy (FCMs)	O algoritmo é rotulado como dMAGAFCM - GRN, onde agentes e seus comportamentos são projetados com propriedades intrínsecas aos problemas de reconstrução GRN.	Os resultados experimentais mostram que o dMAGAFCM -GRN é capaz de aprender FCMs até 200 nós de maneira eficaz. Para isso, 40.000 pesos são otimizados.
[Khan et al., 2003]	Descreve a simulação de redes de transdução de sinal (ST) usando a arquitetura DECAF MAS.	Os agentes mantêm representações de estado interno da concentração de reagentes; é vital modelar o domínio ST. Para cada agente, há um arquivo de regras com comunicação entre agentes.	O caminho do fator de crescimento epidérmico (EGF) é modelado, mostrando a viabilidade da tecnologia MAS para simular redes biológicas.
[Haydarlou et al., 2016]	Introduz uma estrutura de simulação genérica para o BioPAX.	Ele usa o evento discreto e Sistemas Multiagente. É capaz de gerar automaticamente um Sistema Multiagente hierárquico para um modelo BioPAX específico.	Todas as simulações foram consistentes com os modelos originais.
[Ghazikhani et al., 2011]	Propor um novo algoritmo para inferência de rede regulatória genética	Utiliza Sistemas Multiagente para treinar uma rede neural recorrente (RNN)	Eles comparam seu algoritmo com um algoritmo semelhante que usa RNN e otimização de enxame de partículas (PSO) com o conjunto de dados E. coli SOS.

Tabela 5: Resultado da SLR - Comparação entre artigos - Parte II

Artigo	Objetivo	Técnicas	Resultados
[Sanfilippo et al., 2012]	Apresentar uma abordagem em que a modelagem de emergência seja estendida às redes regulatórias e demonstre sua aplicação na descoberta de vias neuro-protetoras.	O foco na emergência da modelagem é de particular interesse, pois apóia a descoberta de novas vias de rede que emergem iterativamente das propriedades auto-organizacionais de genes e grupos de genes, em oposição a outras abordagens de modelagem dinâmica, como redes bayesianas dinâmicas e dinâmica de sistemas, onde a estrutura da rede permanece inalterada durante o processo de simulação.	Uma avaliação inicial da abordagem indica que a modelagem de emergência fornece novas ideias para a análise de redes regulatórias que podem avançar na descoberta de tratamentos agudos para acidente vascular cerebral e outras doenças.
[Pham, 2008]	Propõe uma abordagem baseada em vários agentes para simular grandes redes booleanas aleatórias.	Cada nó da rede booleana aleatória é modelado como um agente. Os agentes mudam passo a passo, baseando-se no estado de todos os vizinhos.	Foi proposta de uma abordagem simples baseada em vários agentes para implementar a simulação em larga escala de redes booleanas aleatórias. Para avaliar, foi apresentada a análise teórica da dependência entre o desempenho e a razão de computação para comunicação.
[Yang and Sun, 2011]	Apresentar um novo método para reconstruir a rede usando Sistemas Multi-agente para fundir os dados de expressão gênica e os dados de ligação ao Fator de Transcrição (TF).	Baseando-se na fusão, uma rede inicial é gerada. Para obter a rede final, é utilizado o método de rede dinâmica bayesiana (DBN).	Os experimentos foram com 25 genes e os resultados foram comparados com os algoritmos XX. Usando MAS e DBN, os resultados tiveram um melhor desempenho.

3.1 Descrição dos trabalhos relacionados resultantes da SLR

No artigo [Liu et al., 2016] foi desenvolvido um algoritmo genético dinâmico multiagente (dMAGA) afim de reconstruir GRN. A proposta é para reconstruir GRN a partir de perfis de expressão de séries temporais baseado em mapas cognitivos fuzzy (FCMs). Este algoritmo foi denominado (dMAGAFCM-GRN). Neste projeto todos os agentes e seus comportamentos são projetados com as propriedades intrínsecas de problemas de reconstrução GRN. Todos os agentes vivem em um ambiente tipo treliça, e os vizinhos de cada agente mudam dinamicamente de acordo com sua energia em cada geração. Na comparação com outros cinco algoritmos existentes percebe-se que dMAGAFCMGRN supera todos os outros algoritmos e pode aproximar a série temporal com alta precisão.

No trabalho desenvolvido em [Khan et al., 2003] foi aplicado a plataforma do sistema multiagente (MAS) a tarefa de simulação de redes biológicas. E foi escrito a simulação de redes de transdução de sinal (ST) usando a arquitetura DECAF MAS. Ao contrário de abordagens anteriores que dependiam de sistemas de equações diferenciais (DE), o framework distribuído do MAS escala bem e permite modelar grandes vias ST altamente interconectadas. Esta escalabilidade é conseguida através da adoção de uma estratégia híbrida, tais como taxa de reação “stants”, para calcular a cinética estocástica no nível de moléculas individuais. Assim, capturando o domínio ST em um nível intermediário de abstração.

No trabalho [Haydarlou et al., 2016], o BioPAX foi definido como uma linguagem padrão baseada na web semântica para especificar modelos de rede biológica a nível molecular e celular. No entanto, uma estrutura de software genérica para simular modelos BioPAX não existe. Para suprir essa ausência foi criado o BioASF (BioPAX-based Agent-oriented Simulation Framework), um framework que separa explicitamente o modelo de execução da estrutura do modelo como fornecido pelo BioPAX, com a vantagem de que o processo de modelagem se torna mais reproduzível e intrinsecamente mais modular; isso garante que as restrições biológicas naturais sejam satisfeitas na execução. A estrutura é baseada nos princípios de sistemas de eventos discretos e sistemas multiagente, e é capaz de gerar automaticamente um sistema multiagente hierárquico para um determinado modelo BioPAX.

O Framework foi desenvolvido em Java e segundo os autores obteve resultados consistentes com os resultados de simulação relatados pelos pesquisadores que desenvolveram os modelos originais [Haydarlou et al., 2016].

O BioASF usa uma linguagem padrão (BioPAX) para especificar os modelos de vias regulatórias. No BioASF, os nós da rede têm um significado semântico natural, são organizados hierarquicamente e a rede exibe um alto grau de modularidade. BioASF recebe a sua modularidade da BioPAX, em que as entidades biológicas são definidas de forma hierárquica, proporcionando modularidade natural para o modelo [Haydarlou et al., 2016].

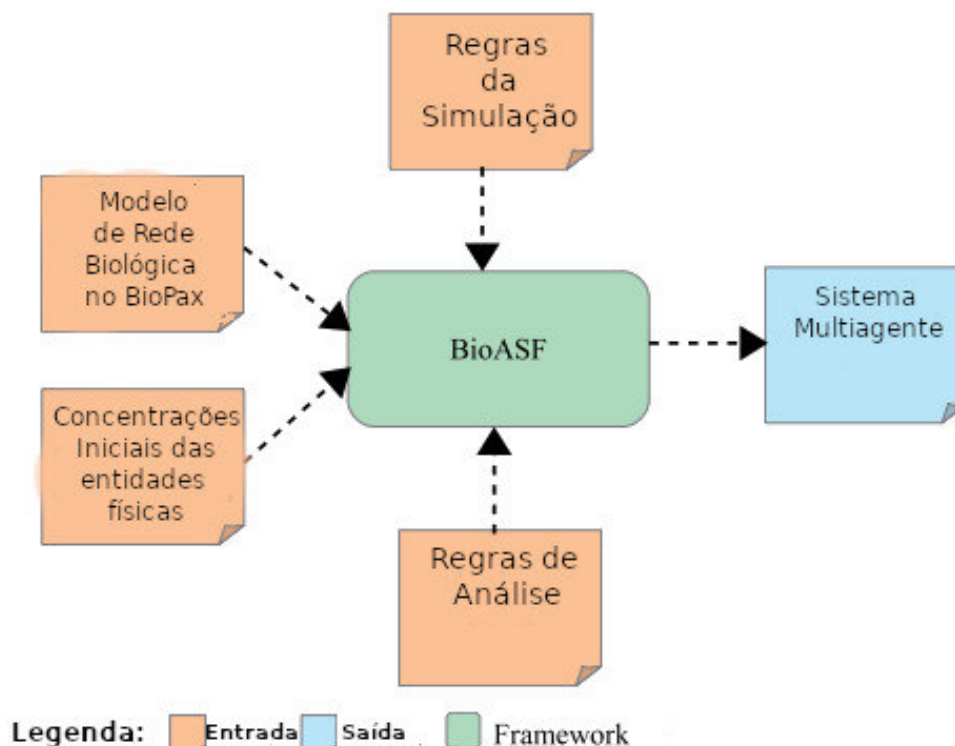


Figura 3.4: Adaptada de [Haydarlou et al., 2016]: O modelo tem como entrada o modelo de rede biológica no BioPAX, parâmetros físicos iniciais, regras da simulação e regras de análise. A saída é um sistema multiagente que é gerado automaticamente por BioASF e é responsável pela execução de interações BioPAX

O trabalho desenvolvido está estruturado na Figura 3.4. Para validar o BioASF, foram utilizados dois modelos diferentes de redes biológicas no BioASF: uma rede de regulação genética modelando os reguladores de células-tronco hematopoiéticas (HSC) e uma rede de transdução de sinal que modela a via de sinalização Wnt / b-catenina. Ambos os modelos originais foram desenvolvidos em Redes de Petri. Observamos que os resultados das simulações realizadas com o BioASF foram semelhantes aos resultados de validação [Haydarlou et al., 2016].

O framework BioASF pode ser usado em diferentes cenários. Para por exemplo, a estrutura pode ser usada para calibração do modelo. Neste cenário, um modelo BioPAX pode ser visualizado no SBGN e mostrado para usuários. Os usuários podem definir break-points e watch-points em caminhos e interações, semelhantes à depuração de códigos de aplicativos. Sendo assim, este modelo apresentou bons resultados, porém é híbrido [Haydarlou et al., 2016].

O artigo [Ghazikhani et al., 2011] foca no fato de que as GRN são não-lineares, dinâmicas e ruidosas. Dados de séries temporais têm sido frequentemente usados para modelagem GRN. Devido a função de aproximação e natureza de feedback da GRN, um modelo de Rede Neural Recorrente (RNN) é usado. O modelo de confiança, negociação e comunicação (Trust, Negotiation and Communication - TNC) para a arquitetura MAS baseia-se no conceito de confiança. Agentes interagem uns com os outros de acordo com a confiança que eles têm uns nos outros. No modelo TNC, a confiança é conceituada como um peça de informação que é trocada entre os agentes no MAS.

De acordo com o trabalho desenvolvido em [Yang and Sun, 2011], técnicas como modelagem baseada em agentes e simulações multiagente são de particular interesse, pois apóiam a descoberta de caminhos emergentes, ao contrário de outras abordagens dinâmicas de modelagem, como redes Bayesianas dinâmicas e dinâmica do sistema. Foram desenvolvidos modelos baseados em agentes (ABM) para dois tratamentos medicamentosos nos dados LPS e CpG, utilizando os clusters e ODEs e, em seguida, calculou o coeficiente de correlação entre simulações dos modelos e os dados observados. O NetLogo foi utilizado para implementar o modelo baseado em agente que simula a dinâmica neuroprotetora da rede de regulação pós-Acidente Vascular Cerebral.

A proposta do trabalho apresentado em [Pham, 2008] é desenvolver um sistema multiagente baseado numa abordagem para simular grandes e aleatórias redes booleanas. Modelou-se a Rede Booleana Aleatória como um sistema multiagente. Cada nó da rede é modelado como um agente que consiste em três campos principais: estado do agente que é uma variável booleana, lista de seus vizinhos, que é um array dinâmico (porque as redes podem ser não homogêneas, k não é o mesmo para todos os agentes, como uma rede sem escala tem uma distribuição exponencial de conectividade, embora a maioria dos nós possa ter conectividade pequena, alguns podem ter grandes valores) e a tabela de consulta que pode ser representada como uma (matriz de inteiro). Os agentes se comportam passo a passo. Em cada passo de sua evolução, um agente lê o estado de todos os seus vizinhos refere-se à sua tabela de consulta e, em seguida, atualiza seu estado.

O trabalho desenvolvido no artigo [Yang and Sun, 2011] é apresentar um novo método para reconstruir redes. Neste método, é usado um sistema multiagente (MAS) para fundir os dados de expressão gênica e dados de ligação de TF e gerar uma rede inicial. Com base na rede inicial, uma rede final é aprendida usando método de aprendizagem de Rede Bayesiana Dinâmica (DBN). Para verificar o desempenho deste método, foi experimentado o método usando os dados de 25 genes e comparado com o resultado de algoritmos já levantado nos artigos anteriores. O resultado comparativo mostra que o método baseado em MAS e DBN tem um melhor desempenho do que outros.

Um dos trabalhos apresentados na tabela 4 merece uma atenção e explicação mais completa. Trata-se do trabalho [Alon, 2007], que trata dos “motifs” de uma rede como já explicado anteriormente. Segundo o autor, as redes regulatórias de transcrição controlam a expressão de genes. As redes de transcrição de microrganismos bem estudados parecem ser constituídas por um pequeno conjunto de padrões de regulação recorrentes, chamados “motifs” de rede [Alon, 2007].

No “C1-FFL” (um dos tipos de “motifs” descritos em [Alon, 2007] e já mencionados neste trabalho), X e Y representam os genes ativadores do processo de transcrição [Alon, 2007]. Aqui será considerada apenas uma função de entrada “AND” para fins de comparação com trabalho que está sendo desenvolvido.

Com uma função de entrada “AND”, o “C1-FFL” mostra um atraso após a estimulação, mas não demora quando a estimulação para. Para ver isso, seguir-se-á o comportamento do “C1-FFL” [Alon, 2007].

Quando o sinal S_x (fator externo) aparece, X fica ativo e liga rapidamente seus promotores. Como um resultado, Y começa a acumular-se. No entanto, devido à Função de entrada “AND”, a produção Z começa apenas quando a concentração de Y ultrapassa o limiar de ativação para o promotor Z. Isso

resulta em um atraso na expressão de Z após o aparecimento de Sx. Por outro lado, quando o sinal Sx é removido, X rapidamente se torna inativo [Alon, 2007].

Como resultado, parar a produção de Z requer que apenas um braço da porta “AND” seja “desligado”. Portanto, não há atraso na desativação de Z após a remoção do sinal Sx [Alon, 2007].

Esse comportamento dinâmico é chamado de atraso sensível ao sinal, isto é, o atraso depende do sinal da etapa Sx. O passo “ON” (adição de Sx) causa um atraso na expressão Z, mas uma etapa “OFF” (remoção de Sx) não causa atrasos. A duração do atraso é determinada pelos parâmetros bioquímicos do regulador Y. Tal comportamento está representado nos gráficos da Figura 3.5.

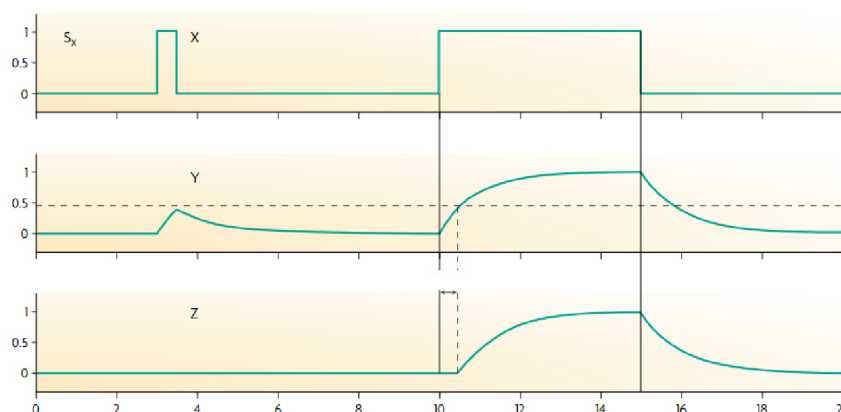


Figura 3.5: Comportamento da simulação utilizando uma porta AND [Alon, 2007]

Porém em [Alon, 2007] não é apresentado maiores detalhes da calibração e do modo como foram gerados os gráficos, sendo assim dificultando a replicação dos seus resultados de forma mais exata.

3.2 Revisão Sistemática Complementar

Na busca de uma revisão sistemática que abrangesse trabalhos publicados até o ano de 2021, foi realizada uma revisão sistemática complementar.

O protocolo seguido é o mesmo descrito na revisão sistemática realizada inicialmente, baseado na padronização descrita neste guia [Mariano et al., 2017].

As bases de dados definidas foram: Pubmed¹⁰, PMC¹¹, Science Direct¹², Scopus¹³ e Scielo¹⁴ e as palavras-chave definidas foram: “regulatory network” AND “multiagent system”. Porém, estas palavras-chave podem ter variantes na escrita e as buscas nas bases de dados são realizadas de maneiras diferentes, podendo variar também os operadores relacionais e expressões regulares [Agostinho et al., 2020a].

As Tabelas 6, 7, 8 e 9 possuem as relações de palavras-chaves e suas variações utilizadas em cada base dados. Esta busca foi limitada ao período de 2018 à 2021 para complementar a revisão

¹⁰<https://www.ncbi.nlm.nih.gov/pubmed/>

¹¹<https://www.ncbi.nlm.nih.gov/pmc/>

¹²<https://www.sciencedirect.com/>

¹³<https://www.scopus.com/home.uri>

¹⁴<https://www.scielo.org/>

sistemática inicial.

Tabela 6: Bases de dados e todas as variantes de palavras-chave - PUBMED

Palavras-chave - PUBMED
regulatory network AND multiagents systems
regulatory network AND multiagents
regulatory network AND multiagents system
regulatory network AND multiagent system
regulatory network AND multiagent systems
“regulatory network AND multiagents systems”
“regulatory network AND multiagents”
“regulatory network AND multiagents system”
“regulatory network AND multiagent system”
“regulatory network AND multiagent systems”
regulatory pathways AND multiagent
regulatory pathways AND multiagents
regulatory Pathways AND multiagents system
regulatory pathways AND Multiagent System
regulatory pathways AND multiagent systems
regulatory network AND multi-agent
regulatory network AND multi-agents systems
regulatory network AND multi-agents
regulatory network AND multi-agents system
regulatory network AND multi-agent system
regulatory network AND multiagent systems
regulatory pathways AND multi-agent
regulatory pathways AND multi-agents
regulatory Pathways AND multi-agents system
regulatory pathways AND Multi-agent System
regulatory pathways AND multi-agent systems
regulatory network AND multi-agent

Tabela 7: Bases de dados e todas as variantes de palavras-chave - PMC

Palavras-chave - PMC
regulatory network AND multiagents systems
regulatory network AND multiagents
regulatory network AND multiagents system
regulatory network AND multiagent system
regulatory network AND multiagent systems
regulatory pathways AND multiagent
regulatory pathways AND multiagents
regulatory Pathways AND multiagents system
regulatory pathways AND Multiagent System
regulatory pathways AND multiagent systems
regulatory pathways AND multiagent systems
regulatory network AND multiagent
regulatory network AND multi-agents systems
regulatory network AND multi-agents
regulatory network AND multi-agents system
regulatory network AND multi-agent system
regulatory network AND multi-agent systems
regulatory pathways AND multi-agent
regulatory pathways AND multi-agents
regulatory pathways AND Multi-agent System
regulatory pathways AND multi-agent systems
regulatory network AND multi-agent

Tabela 8: Bases de dados e todas as variantes de palavras-chave - Science Direct

Palavras-chave - Science Direct
regulatory network AND multiagents systems
regulatory network AND multiagents
regulatory network AND multiagents system
regulatory network AND multiagent system
regulatory network AND multiagent systems
regulatory pathways AND multiagent
regulatory pathways AND multiagents
regulatory Pathways AND multiagents system
regulatory pathways AND Multiagent System
regulatory pathways AND multiagent systems
regulatory network AND multiagent
regulatory network AND "multi-agents"systems
regulatory network AND "multi-agents"
regulatory network AND "multi-agents"system
regulatory network AND "multi-agent"system
regulatory network AND "multi-agent"
regulatory network AND "multi-agent"
regulatory pathways AND "multi-agent"
regulatory pathways AND "multi-agents"
regulatory Pathways AND "multi-agents"system
regulatory pathways AND "Multi-agent"System
regulatory pathways AND "multi-agent"systems
regulatory network AND "multi-agent"

Tabela 9: Bases de dados e todas as variantes de palavras-chave - Scopus

Palavras-chave - Scopus
“regulatory network”AND ”multiagents systems”
“regulatory network”AND ”multiagents”
“regulatory network”AND ”multiagents system”
“regulatory network”AND ”multiagent system”
“regulatory network”AND ”multiagent systems”
“regulatory pathways”AND ”multiagent”
“regulatory pathways”AND ”multiagents”
“regulatory Pathways”AND ”multiagents system”
“regulatory pathways” AND “Multiagent System”
“regulatory pathways” AND “multiagent systems”
“regulatory network” AND “multiagent”
“regulatory network” AND “multi-agents systems”
“regulatory network” AND “multi-agents”
“regulatory network” AND “multi-agents system”
“regulatory network” AND “multi-agent system”
“regulatory network” AND “multi-agent systems”
“regulatory pathways” AND “multi-agent”
“regulatory pathways” AND “multi-agents”
“regulatory Pathways” AND “multi-agents system”
“regulatory pathways” AND “Multi-agent System”
“regulatory pathways” AND “multi-agent systems”
“regulatory network” AND “multi-agent”

As bases de dados utilizadas apresentaram mudanças na sintaxe de busca e devido a este fato foram introduzidas mais variantes das palavras-chave visando realizar uma busca mais abrangente. Antes da etapa de eliminação de artigos duplicados obteve-se um total de 10.096 artigos, essa etapa reduziu para 2.656 artigos. Foram seguidos os passos descritos no guia [Mariano et al., 2017] e então obteve-se o seguinte resultado descrito na Tabela 10.

Tabela 10: Número de artigos em cada etapa

Etapa	Número de artigos
Após eliminação de duplicados	2.656
Filtro por Título	15
Filtro por resumo	6
Filtro pela leitura Diagonal	1
Filtro pela leitura completa	1
Resultados	1

O artigo resultante foi “Agent-based modeling of morphogenetic systems: Advantages and challenges” [Glen et al., 2019]. Este artigo trata-se de uma Revisão onde são destacados os benefícios e desafios técnicos associados aos Modelos baseados em agente como ferramentas para examinar eventos morfogenéticos.

4 METODOLOGIA

A proposta deste trabalho envolve o desenvolvimento de um ambiente de simulação multiagente para vias regulatórias genéticas que seja capaz de representar GRN com MAS, diminuindo a necessidade de uma grande quantidade de parâmetros e tornando o processo mais flexível representando cada indivíduo presente no sistema. Para tanto, a modelagem é toda proposta na ideia de “Motifs” como estrutura básica.

O modelo proposto é uma abstração dos processos bioquímicos subjacentes e simula a interação entre os TFs e um gene alvo. Sendo as ferramentas MAS, muito flexíveis, apresenta-se uma versão inicial de um MAS como “framework” de modelagem para representar uma rede regulatória [Agostinho et al., 2019].

O principal componente nesse modelo é o Sítio de Ligação, que é sensível a TFs nas proximidades, dependendo da quantidade de TFs reguladores. Quando essa quantidade de TFs reguladores definida no arquivo de atributos de entrada estiver no Sítio de Ligação, a transcrição (ativação ou inibição) está ativada [Agostinho et al., 2019].

O simulador é implementado em NetLogo e são definidos dois arquivos de configuração para o modelo: definição de agentes (atributos) e definição de restrições (Tabelas 11 e 12) [Agostinho et al., 2019].

A ferramenta foi desenvolvida em três etapas, sendo que na primeira tem os resultados validados através do processo de Michalis-Menten; a segunda etapa tem seu desenvolvimento evoluído e sua validação é realizada através dos resultados do trabalho de Uri Alon; e por último, a etapa final evolui para desenvolvimento final da ferramenta e é validada pelo estudo de caso da planta *Arabidopsis Thaliana*, porém é capaz de simular GRNs variadas sendo configurados corretamente os arquivos de entrada para configurações iniciais.

Tabela 11: Definição dos agentes

Id	Tam	Cor	Qtd	Gene
int	int	cor	int	nome do gene
int	int	cor	int	nome do gene

As configurações contidas na Tabela 11 definem cada gene em termos do tamanho (num plano bidimensional) da primeira etapa do desenvolvimento do simulador. Neste arquivo Id é um inteiro que representa o identificador do gene, cor é qual cor que será atribuída para o agente que representa o gene e o nome do gene.

Tabela 12: Definição das restrições

Regulador	Ação	Regulado	Qtd
nome do gene regulador	booleano	“X”	inteiro

As restrições definidas na Tabela 12 descrevem a relação regulatória entre um regulador e regulado (nome), a ação (ativação ou inibição do processo de transcrição - booleano), e a quantidade (inteiro)

de reguladores necessários para a ação a ser executada. Possuindo essas informações é possível criar o modelo que simulará a GRN [Agostinho et al., 2019]. Esta versão é validada ao obter comportamentos semelhantes aos comportamentos apresentados em [Johnson and Goody, 2011] e descritos na subseção 2.4.1.

Por exemplo, considerando os dados das Tabelas 11 e 12, a GRN resultante terá dois agentes, um regulador (Y) e um regulado (X). A ação de Y sobre X é de ativação (Ação 1) e 3 agentes de Y são necessários para produzir um agente do tipo X. No ambiente do simulador, X e Y são representados, respectivamente, como círculos brancos e vermelhos.

Uma área é definida para representar o Sítio de Ligação no ambiente. Quando um agente Y entra no Sítio de Ligação, ele irá parar de se mover e inicia a sua taxa de decaimento que é necessária ser incluída na simulação, pois imita comportamento real dos TFs, semelhante a uma perda de energia [Alon, 2007]. Simultaneamente, os agentes do tipo X são produzidos, se pelo menos 3 agentes de Y estão presentes no Sítio de Ligação.

Utilizou-se as configurações de entrada como mostradas nas Tabelas 11 e 12. Na Figura 4.1, apresenta-se a versão inicial que ainda possui alguns parâmetros configurados na interface [Agostinho et al., 2020b].

Apresenta-se na Figura 4.1 os parâmetros S_x e S_y que correspondem a fatores externos que podem alterar o funcionamento do processo no Sítio de Ligação, pois depende desses fatores estar habilitada a reação que ocorre no Sítio de Ligação entre os TFs. Os parâmetros $rate1$ e $rate2$ atuam como taxa de decaimento máxima. Já os parâmetros $die-rate1$ e $die-rate2$ atuam como as taxas de decaimento nesta versão. O restante dos visualizadores mostra informações sobre a quantidade de agentes e sobre o decaimento [Agostinho et al., 2020b].

Código 1: Controle das taxas de decaimento

```
while [ member? protein_name list_rest and pcolor = orange ][

    if ((decay1 < die_rate1) and regulator = 1)[
        die
        set tam1 tam1 - 1
    ]
    if ((decay2 < die_rate2) and regulator = 1)[
        die
        set tam2 tam2 - 1
    ]
    if ((decay1 <= 1) and regulator = 1)[
        set decay1 (decay1 - rate1)
        set tam1 tam1 - decay1
    ]
    if ((decay2 <= 1) and regulator = 1)[
```

```

    set decay2 (decay2 - rate2)
    set tam2 tam2 - decay2
]
set monitor_decay1 decay1
set monitor_decay2 decay2

```

```
]
```

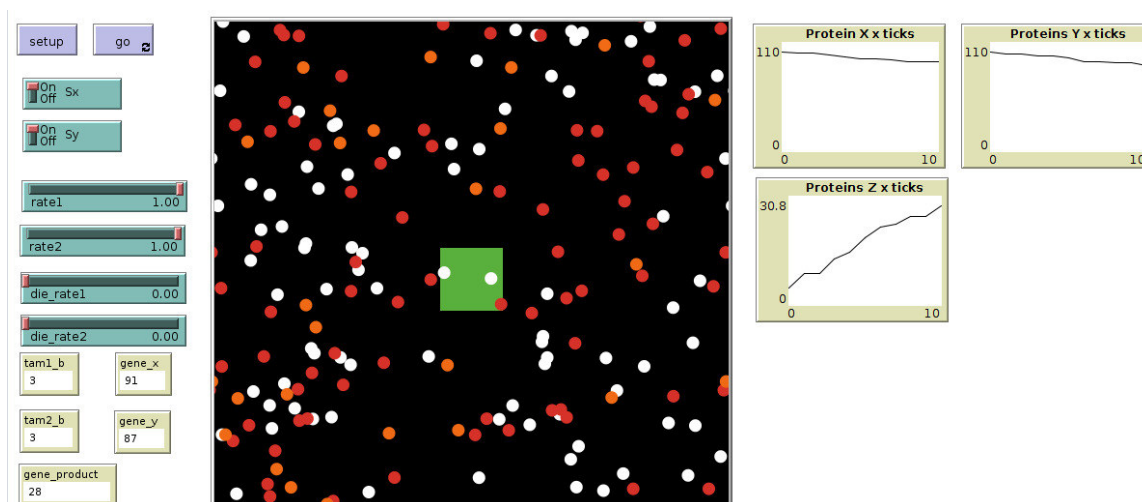


Figura 4.1: Interface do modelo Inicial

O Código 2 contém a função de chamada das funções responsáveis pela configuração de parâmetros e da função responsável pela execução do modelo. Já os Códigos 3 e 4 são responsáveis por, efetivamente, configurar os atributos e restrições para a execução do modelo. E o Código 1 controla a variação das taxas de decaimento [Agostinho et al., 2020b].

Código 2: Função de chamada das funções responsáveis pela configuração de parâmetros e da função responsável pela execução do modelo

```

to import-network
  clear-all
  set-default-shape turtles "circle"
  import-attributes
  import-restrictions
  reset-ticks
end

```

Código 3: Função de configuração dos parâmetros

```

to import-attributes

```

```

file -open "attributes4.txt"
set att table:make

while [not file -at-end?]
[
  let items read-from-string (word "[" file-read-line "]")
  let id (item 0 items)
  table:put att item 4 items items

  set n item 3 items
  ;#####
  ;default configuration file attributes
  ; node-id size    color    qtd      protein_name
  ;  0         1     white    50      "X"

  create-proteins n [
    setxy random-xcor random-ycor
    set node-id item 0 items
    set size    item 1 items
    set color    item 2 items
    set waiting1 0
    set waiting2 0
    set waiting  0
    set protein_name item 4 items
    set regulator 0
  ]
]

file -close
end

```

Código 4: Função de configuração das restrições

```

to import-restrictions

  file -open "restrictions4.txt"
  set rest table:make

```

```

while [not file-at-end?]
[
  let items read-from-string (word "[" file-read-line "]")

  ifelse (empty? items)[

  ]
  [ let id (item 0 items)
    table:put rest id items
  ]

]
file-close
end

```

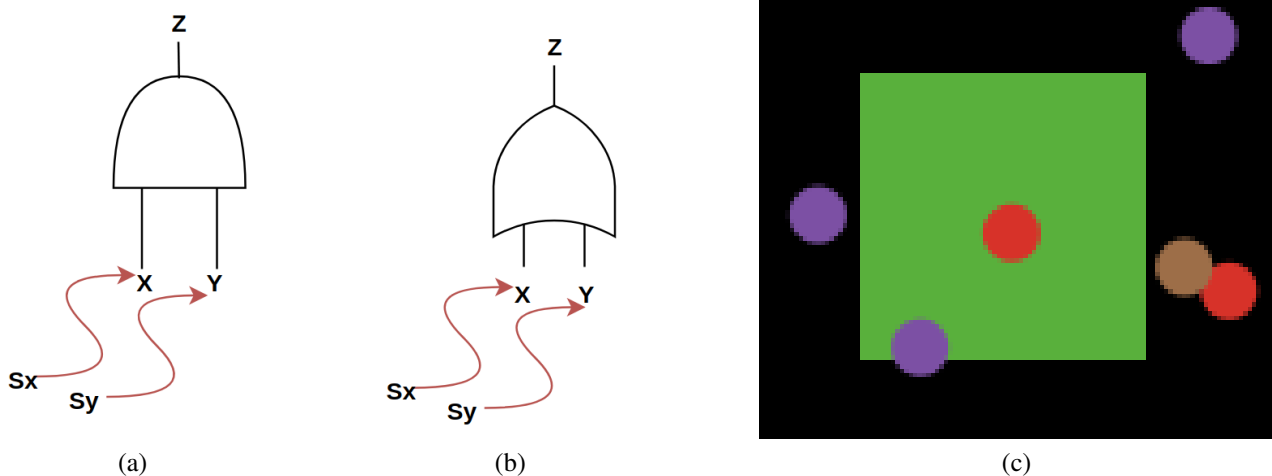
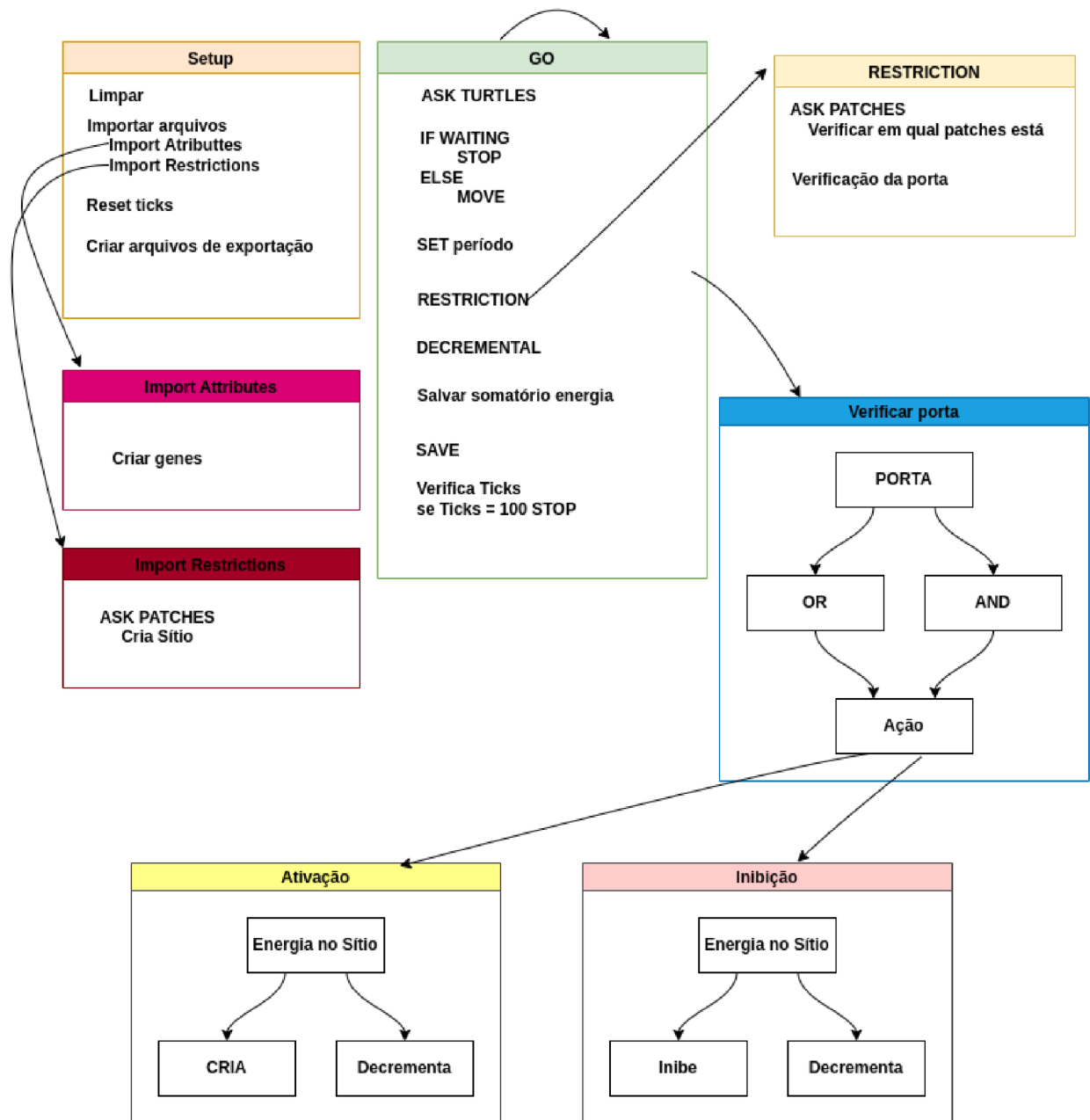


Figura 4.2: (a) “motif C1-FFL” com uma função “AND”. (b) “motif C1-FFL” com uma função “OR”, (c) representação do Sítio de ligação e TFs

t

A Figura 4.2a representa como é o funcionamento do modelo baseado em Motifs com lógica AND e a Figura 4.2b representa o funcionamento do modelo baseado em motifs com porta OR, e em cada uma dessas representações há os fatores externos Sx e Sy que estimulam a ação (produção ou inibição do produto). Já a Figura 4.2c possui a área em verde que representa o Sítio de Ligação, a esfera vermelha representa o TF X, a esfera roxa é o TF Y e a esfera marrom é o produto Z. Assim, na Figura 4.2 tem-se a representação do modelo proposto em sua unidade menor que é um Sítio de Ligação, dois TFs (ainda que possam representar a mesma unidade genética, ou seja, mesmo gene) e um produto que é modificado através da ação de ativação ou inibição. Portanto, há nesta Figura a representação da base do funcionamento do modelo.

Nos arquivos de entrada será preciso definir algumas outras especificações relacionadas a parte visual do desenvolvimento do modelo, como o posicionamento e os tamanhos dos Sítios de Ligação e também especificações como quantos TFs são necessários estar no Sítio de ligação para que possa ocorrer a ação. A metodologia desenvolvida até aqui é a metodologia base simplificada desenvolvida para validação com estudo de caso da teoria de Michaelis e Menten 2.4.1. Posteriormente, será descrita toda a metodologia desenvolvida para o framework, como representado na Figura 4.3.



(a)

Figura 4.3: Funcionamento do Modelo - esquema de funcionamento do framework

O esquema mostrado na Figura 4.3 representa sintetizadamente o funcionamento do framework.

Setup é a função principal de inicialização onde o grid é limpo, após são chamadas as funções para importação de dados que são Import Attributes e Restrictions onde, Import Attributes carrega as informações sobre os atributos necessários para gerar os TFs. Sendo assim, Import Attributes cria as esferas que são os genes no modelo e Restrictions contém as informações de restrições e informações necessárias para a simulação como a quantidade de genes necessários para ocorrer a ação, qual a ação e informações necessárias para criação dos “patches” que são os Sítios de Ligação, além de outras informações que são detalhadas posteriormente. Após, os ticks são inicializados e são criados os arquivos de exportação de informações.

No esquema, em G0 é feita a leitura dos agentes em uma ordem aleatória. Quando essa leitura é feita o agente ou conjunto de agentes especificado executa os comandos fornecidos, ou seja, os genes executarão os comandos especificados que são descritos a seguir. É realizada a verificação se o gene(agente) lido está na região correspondente ao Sítio de Ligação em que ocorre a reação em que ele é TF, se estiver, o agente pára de se movimentar, se não estiver, continua se movimentando aleatoriamente.

Posteriormente é configurado em qual turno o gene (agente) encontra-se, se é “morning”, “central”, “vespertine” ou “nocturnal”, respectivamente se for ($\text{ticks} \leq 25$), ($\text{ticks} > 25$) e ($\text{ticks} \leq 50$), ($\text{ticks} > 50$) e ($\text{ticks} \leq 75$), ($\text{ticks} > 75$) e ($\text{ticks} \leq 100$). A Figura 4.4 representa a divisão do modelo em períodos e os respectivos limites em ticks.

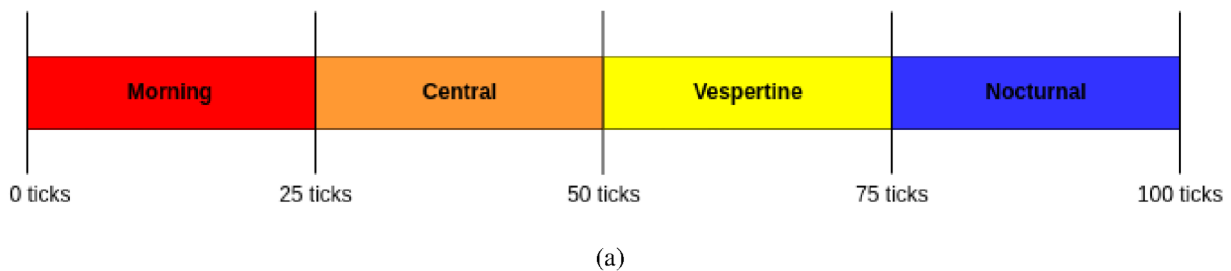


Figura 4.4: Escala que representa os períodos definidos no modelo.

Na etapa posterior é chamada a função RESTRICTION que realiza a leitura dos patches, ou seja, verifica em qual Sítio de ligação ocorrerá as próximas verificações, por exemplo a verificação da porta lógica, o framework irá realizar as ações dependendo da porta configurada “AND” ou “OR” e então se houver energia suficiente nos TFs envolvidos que encontram-se no Sítio de Ligação, ocorrerá a ação que poderá ser de Ativação ou Inibição. Se a ação for de ativação ocorrerá a criação de mais genes (produto), se a ação for de inibição, ocorrerá a exclusão de genes (produto). Ao mesmo tempo que ocorrem essas ações ocorrem decrementações da energia, simulando o gasto de energia dos genes que ocorre na natureza.

5 ESTUDOS DE CASO

Nesta seção são apresentados os estudos de caso aplicados na ferramenta proposta. O primeiro estudo de caso é o da teoria de Michaelis e Menten (descrito na seção 2.4.1), que é processo de regulação entre os genes que emerge das interações nos sítios de ligação e podem ser modelados como um processo de cinética enzimática. O segundo é o Modelo de Uri Alon (descrito na seção 2.4.2), em que os fatores de transcrição que estiverem no sítio de ligação interagem entre si e ativam ou inibem o produto de acordo com porta lógica AND ou OR, e de acordo com a quantidade (energia) desses fatores. Já o último estudo de caso é o Modelo da planta *Arabidopsis Thaliana* (descrito na seção 2.4.3), em que o framework está habilitado para simular a GRN do ciclo da planta ou outras GRNs genéricas que forem configuradas de acordo com os arquivos de entrada.

5.1 Comparação com os resultados da teoria de Michaelis e Menten

A Metodologia desenvolvida para este Estudo de caso é a metodologia simplificada, utilizando Motifs.

5.1.1 Resultados

A definição do agente na Tabela 11 define cada gene em termos de tamanho (em um plano bi-dimensional), cor, quantidade e nome. A definição de restrição da Tabela 12 descreve a relação regulatória entre um regulador e um regulado, sua ação (ativação ou inibição) e a quantidade de reguladores necessários para a ação a ser executada. Com essas informações, é possível criar o modelo que simulará a GRN.

Por exemplo, considerando os dados da Tabela 11 e 12, a GRN resultante terá dois agentes, um regulador (Y) e um regulado (X).

A ação de Y em X é a ativação (Ação 1) e são necessários 3 agentes de Y para produzir um agente do tipo X. No ambiente do simulador, X e Y são representados, respectivamente, como círculos branco e vermelho. Há um quadrado verde no ambiente, que representa o local de ligação para Y. Importante lembrar que no NetLogo, cada etapa do tempo é representada por um “tick”.

Inicialmente, os resultados da simulação não apresentam semelhança com o comportamento da curva resultante do resultado da equação de Michaelis e Menten [Johnson and Goody, 2011]. Mas, pode-se observar que ao decorrer do tempo, com mais “ticks”, a semelhança dos resultados do modelo simplificado com o comportamento da curva de Michaelis e Menten [Johnson and Goody, 2011].

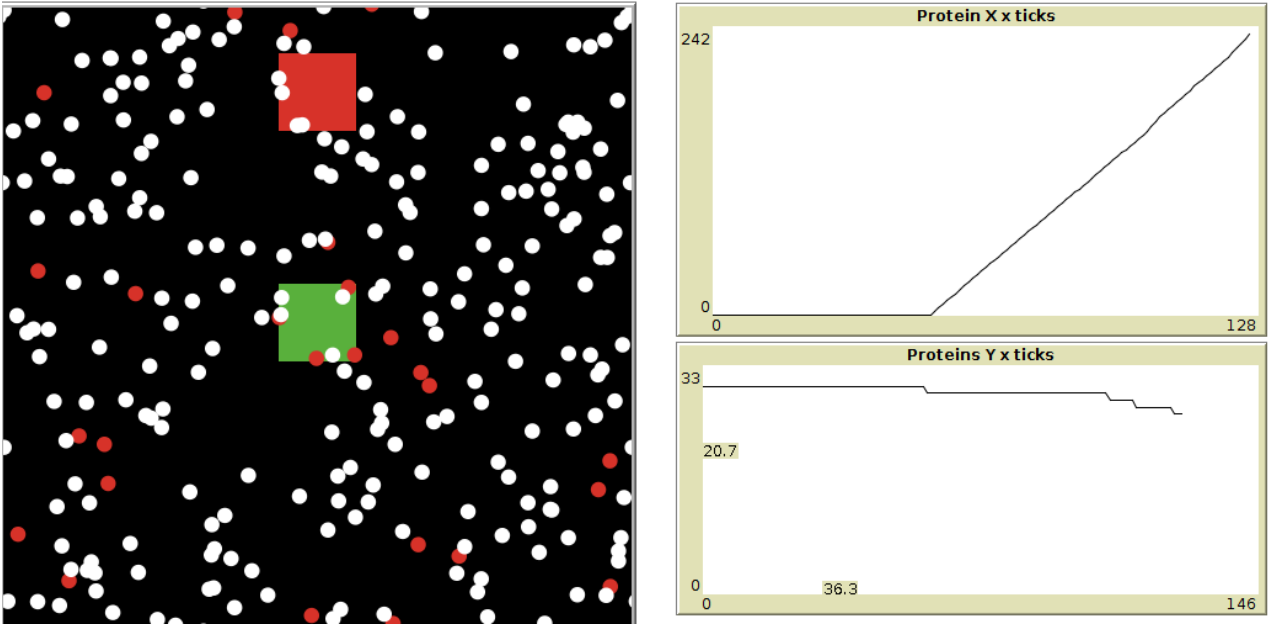


Figura 5.1: Modelo Inicial - 126 ticks

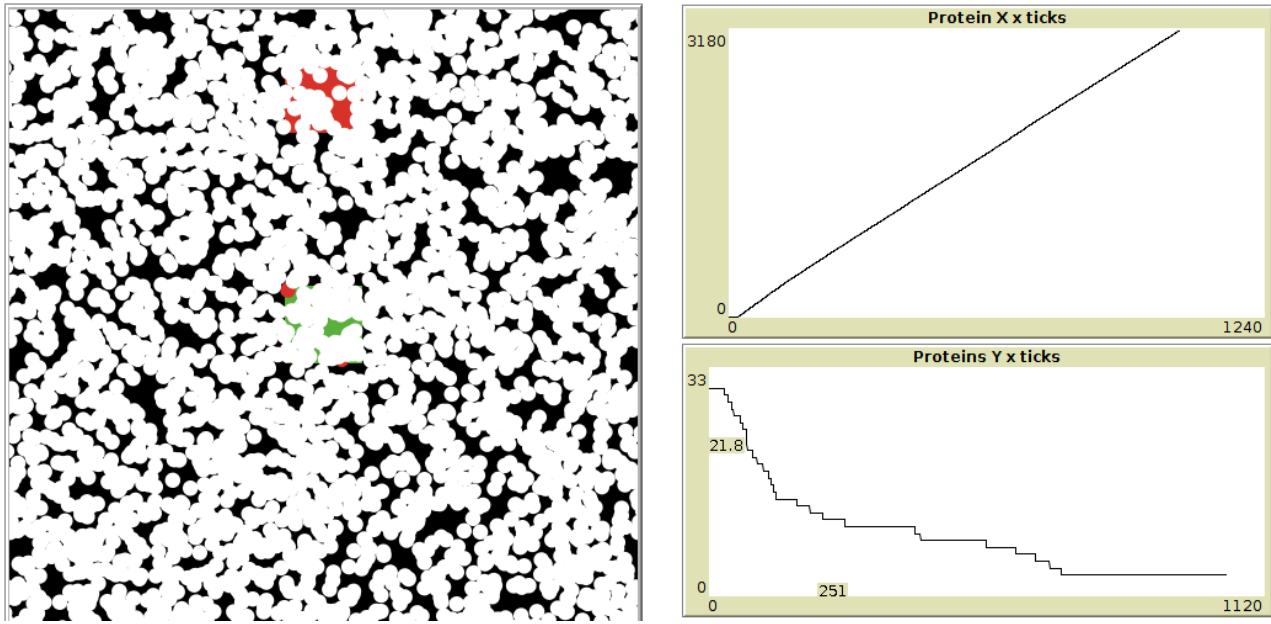


Figura 5.2: Modelo Inicial - 1042 tiks

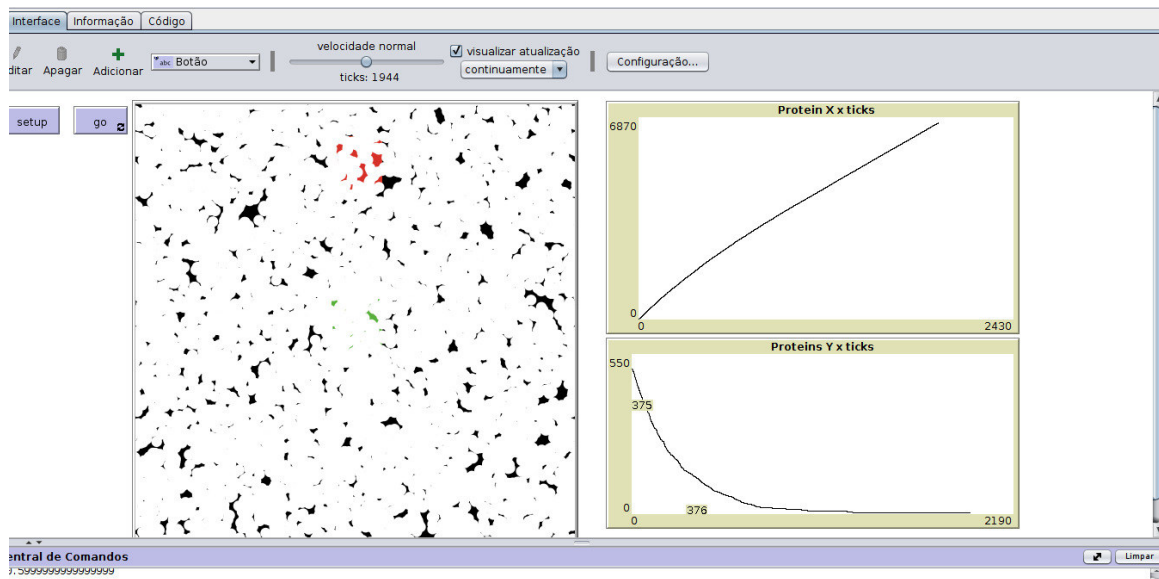


Figura 5.3: Modelo Inicial - 1386 ticks

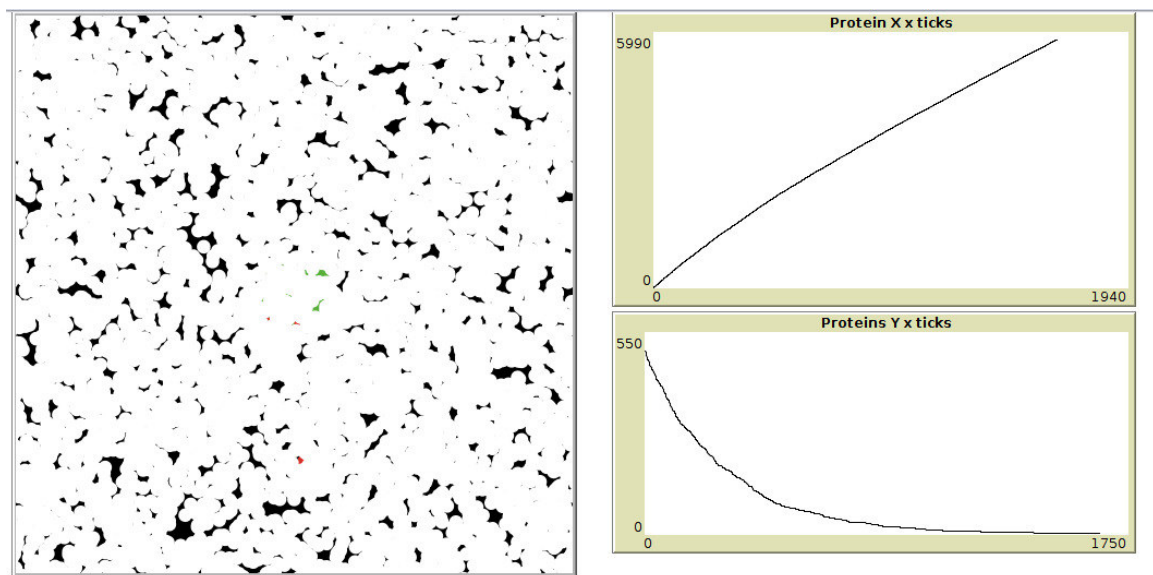


Figura 5.4: Modelo Inicial - 1944 ticks

As Figuras 5.1, 5.2 5.3 e 5.4 representam as simulações realizadas com este modelo, cujo resultados são comparados ao comportamento da curva de Michaelis e Menten [Johnson and Goody, 2011].

A Figura 5.1 mostra a simulação e comportamento alcançado com 126 *ticks*. Pode-se perceber que o comportamento do modelo está bem diferente do esperado.

A Figura 5.2 mostra a simulação e comportamento alcançado com 1.042 *ticks*. Nesta simulação, pode-se perceber que o comportamento está mais aproximado do esperado.

A Figura 5.3 mostra a simulação e comportamento alcançado com 1.386 *ticks*. Nesta simulação, o comportamento está ainda mais aproximado do esperado.

Finalmente, na Figura 5.4 observa-se que o comportamento está bem próximo do comportamento da curva de Michaelis, pois verificas-se que a reação se aproxima assintoticamente de sua taxa máxima

V_{max} [Johnson and Goody, 2011], sendo que o resultado do modelo aproxima-se da curva apresentada na Figura 2.7a que mostra o comportamento da equação de Michaelis-Menten em termos de taxa de reação e a Figura 2.7b mostra as concentrações do substrato e do produto para o mesmo processo.

Todas as simulações foram realizadas com as configurações encontradas nas Tabelas 11 e 12. Obteve-se resultados diferentes pelo fato de que quanto maior o tempo de simulação, há mais possibilidades de ocorrerem mais interações entre os agentes.

5.2 Comparação com o Modelo de Uri Alon

5.2.1 Metodologia

Esta versão é validada ao obter comportamentos como os comportamentos apresentados em [Alon, 2007], ou seja, quando o fator externo está ativado, os fatores de transcrição que estiverem no sítio de ligação interagem entre si e ativam ou inibem o produto de acordo com porta lógica AND ou OR, e de acordo com a quantidade (energia) desses fatores. Esta versão do modelo passa a conter as configurações somente nos arquivos de entrada. Sendo que cada linha dos arquivos de definições representa a configuração de um “Motif” que é componente da GRN a ser simulada.

O modelo proposto é uma abstração dos processos bioquímicos subjacentes e simula a interação entre TFs e um gene alvo. Assim como considerado no modelo inicial, as ferramentas MAS são muito flexíveis permitindo a representação de cada indivíduo em um sistema, simplificando a avaliação de novas hipóteses sobre o modelo. O componente principal também é o sítio de ligação de um gene, que é sensível a TFs próximos. Quando a quantidade definida de TFs regulatórios está no sítio de ligação, a transcrição (ativação ou inibição) é ativada. E para que seja possível realizar comparações com outros modelos, torna-se mais adequado configurar os parâmetros em arquivos de entrada, pois assim espera-se poder replicar configurações de outros modelos.

As Tabelas 13 e 14 apresentam os parâmetros que precisam ser configurados para a simulação ser realizada. Com relação a questão temporal, devido ao fato de não estar explícito em [Alon, 2007], este modelo não apresenta um período de tempo igual ao mostrado em [Alon, 2007].

Tabela 13: Exemplo de arquivo de entrada de definições - Ativação (porta AND) - Comparação Uri Alon

tam	cor	qtd	nome-proteína	nome-regulado	patche-cor-att	porta
int	cor	int	nome-proteína	nome-regulado	cor do patche	booleano
int	cor	int	nome-proteína	nome-regulado	cor do patche	booleano
int	cor	int	nome-proteína	-1	cor do patche	booleano

Na Tabela 13 pode-se verificar os seguintes parâmetros:

- id : identificação do nodo (gene);
- tam: tamanho do agente (gene);
- cor: cor utilizada para distinguir os genes;

- qtd: quantidade inicial de cada gene;
- nome-proteína: nome dado a cada gene;
- nome-regulado : nome do gene regulado (-1 indica que é o próprio regulado)
- “patche-cor-att”: cor do patche (sítio de ligação);
- porta: porta lógica AND ou OR;

Tabela 14: Exemplo de arquivo de entrada de restrições - Ativação (porta AND) - Comparação Uri Alon

id	TF-1	TF-1-v	n-n1	TF-2	TF-2-v	n-n2	regulado	p-cor	px	py
int	nome	bool	[-1, +int]	nome	bool	[-1, +int]	regulado	-1	[-1, +int]	[-1, +int]
int	nome	bool	[-1, +int]	nome	bool	[-1, +int]	regulado	-1	[-1, +int]	[-1, +int]
int	nome	bool	[-1, +int]	nome	bool	[-1, +int]	[-1, +int]	cor	[-int, +int]	[-int, +int]

Na Tabela 14 pode-se verificar os seguintes parâmetros:

- id: identificação do nodo (gene);
- TF-1: nome do TF-1;
- TF-1-v: Se a ação é de ativação o valor é 1, se for de inibição o valor 0;
- n-n1: quantidade do FT1; necessária para ocorrer a produção ou inibição da proteína;
- TF-2: nome do FT2;
- TF-2-v: Se a ação é de ativação o valor é 1, se for de inibição o valor 0;
- n-n2: quantidade do FT2 necessária para ocorrer a produção ou inibição da proteína;
- regulado: qual gene este FT regula;
- “p-cor”: cor do “patche”(sítio de ligação) em que ocorre essa reação;
- px: medida do sítio de ligação ”patche”;
- py: medida do sítio de ligação ”patche”.

5.2.2 Resultados

Nesta versão foram realizados experimentos com a porta AND e função de ativação com 10 rodadas cada um, ou seja, ativa o aumento do produto quando o fator externo estiver ativado e a energia de TF1 e TF2 no sítio de ligação possuir no mínimo o valor especificado no arquivo de entrada de restrições. Também foram realizados três experimentos com a porta AND e função de inibição com 10 rodadas cada um, sendo que o produto é inibido quando o fator externo estiver ativado, e assim como no experimento anterior, é preciso que no sítio de ligação haja energia de TF1 e energia de TF2 com valores iguais ou maiores aos valores necessários especificados no arquivo de entrada.

As Tabelas 15 e 16 apresentam as configurações de entrada para simulações de ativação com porta AND. Os parâmetros “tam”, “cor”, “patche-cor-att” são necessários para o netlogo criar o ambiente, “tam” indica o tamanho do agente, “cor” indica as cores que cada agente apresentará, e “patche-cor-att” define a cor do “patche” que representa o sítio de ligação.

A Tabela 15 possui os parâmetros “qtd”, “nome-proteína”, “regulado” e “porta” que são parâmetros essenciais para a simulação. Sendo assim, pode-se observar que para realizar essa simulação foi definido no parâmetro “qtd” a quantidade de TF necessários para inicializar a simulação, onde tem-se inicialmente 50 agentes que representam TF1, 50 agentes que representam TF2, 6 agentes que representam o produto. A porta lógica configurada para esta simulação é representada pelo parâmetro “porta” com o valor 1 significando a definição da porta AND. TF1 é definido como “X”, TF2 é definido como “Y” e o produto é “Z”. O parâmetro “regulado” definido como “Z” para TF1 e TF2 indica que “Z” é o gene regulado e que nesta simulação ele não regula nenhum gene.

Tabela 15: Arquivo de entrada de definições - Ativação (porta AND) - Resultados - Comparação Uri Alon

tam	cor	“qtd”	nome-proteína	nome-regulado	patche-cor-att	porta
1	violet	50	“X”	“Z”	orange	1
1	red	50	“Y”	“Z”	orange	1
1	brown	6	“Z”	-1	orange	1

A Tabela 16 possui os parâmetros denominados como restrições para a simulação. Estes parâmetros são representados por “id” sendo o agente a ser configurado como gene que interage na simulação, “ft-1” representa o TF1, “ft-1-valor” indica se esse gene é regulado, “n-n1” indica a quantidade de agentes TF1 são necessários para ocorrer uma interação que gere um produto, “ft-2” representa o TF2, “ft-2-valor” indica se esse gene é regulado, “n-n2” indica a quantidade de agentes TF2 são necessários para ocorrer uma interação que gere um produto, regulado define qual gene é o regulado na configuração desse “motif”. Ainda nesta tabela algumas configurações necessárias para a configuração do ambiente no Netlogo que são “patche-cor” que representa em qual “patche” ocorre as interações representadas por esse “motif”. Por último, tempos “px” e “py” que representam coordenadas para o posicionamento o “patche” em que ocorreram as interações. Neste caso, tem-se “X” como TF1, “Y” como TF2, “Z” como regulado, o “patche” definido com a cor verde é o associado a esta configuração e são necessários 1 agente representante de TF1 e um agente representante de TF2 no interior do Sítio de Ligação (“patche”) para o produto “Z” ser gerado se o Fator Externo estiver ativo.

Tabela 16: Arquivo de entrada de restrições - Ativação (porta AND) - Resultados - Comparação Uri Alon

id	ft-1	ft-1-valor	n-n1	ft-2	ft-2-valor	n-n2	regulado	patche-cor	px	py
"X"	"X"	-1	-1	"Y"	-1	-1	"Z"	-1	-1	-1
"Y"	"Y"	-1	-1	"X"	-1	-1	"Z"	-1	-1	-1
"Z"	"X"	1	1	"Y"	1	1	-1	orange	0	0

A Tabela 17 possui a mesma configuração definida na Tabela 15 para os parâmetros, exceto pelo fato de o parâmetro "porta" estar configurado como "0" representando a porta OR.

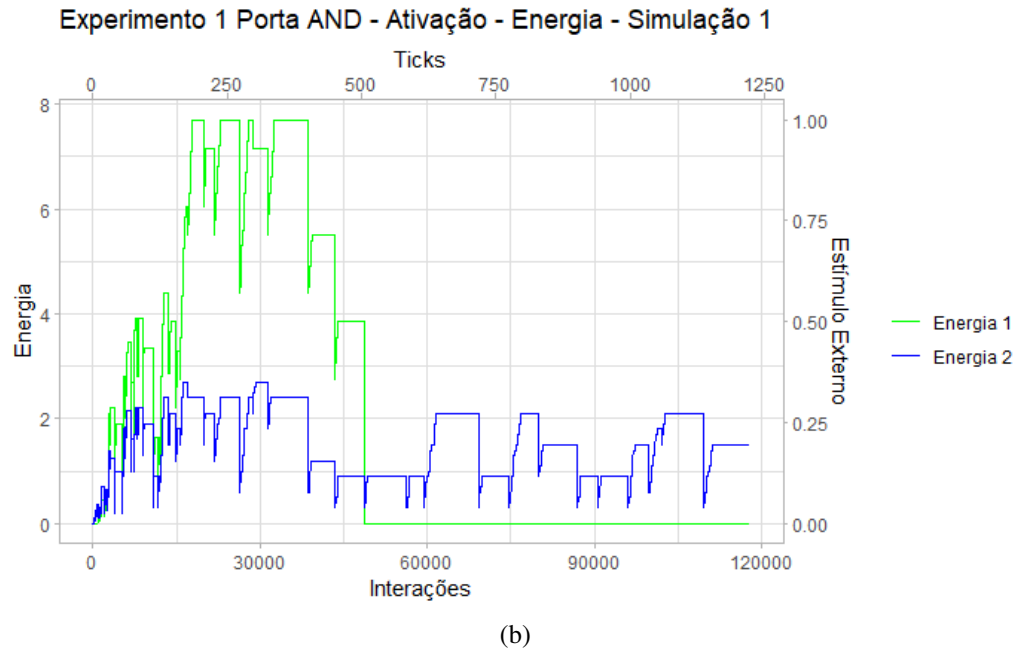
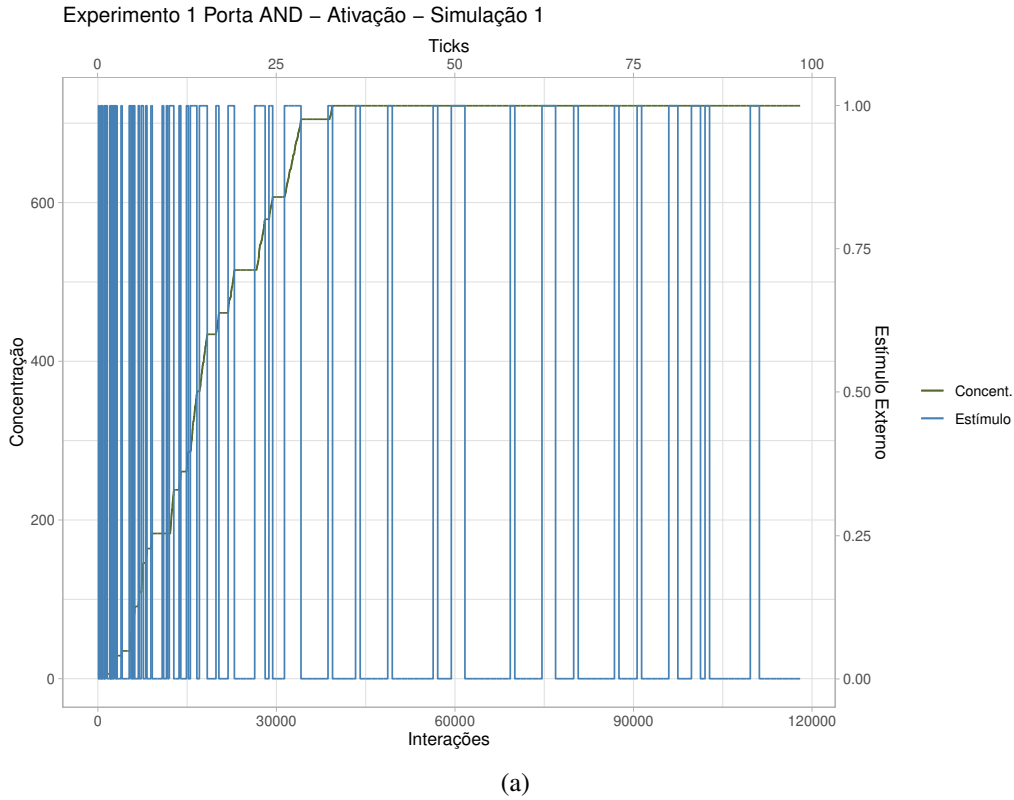
Tabela 17: Arquivo de entrada de definições - Ativação (porta OR) - Resultados - Comparação Uri Alon

tam	cor	qtd	nome-proteina	nome-regulado	patche-cor-att	porta
1	violet	50	"X"	"Z"	orange	0
1	red	50	"Y"	"Z"	orange	0
1	brown	6	"Z"	-1	orange	0

A Tabela 18 possui os mesmos parâmetros de restrições definidos para a simulação de ativação com a porta AND, agora para a porta OR.

Tabela 18: Arquivo de entrada de restrições - Ativação (porta OR) - Resultados - Comparação Uri Alon

id	ft-1	ft-1-valor	n-n1	ft-2	ft-2-valor	n-n2	regulado	patche-cor	px	py
"X"	"X"	-1	-1	"Y"	-1	-1	"Z"	-1	-1	-1
"Y"	"Y"	-1	-1	"X"	-1	-1	"Z"	-1	-1	-1
"Z"	"X"	1	1	"Y"	1	1	-1	orange	0	0



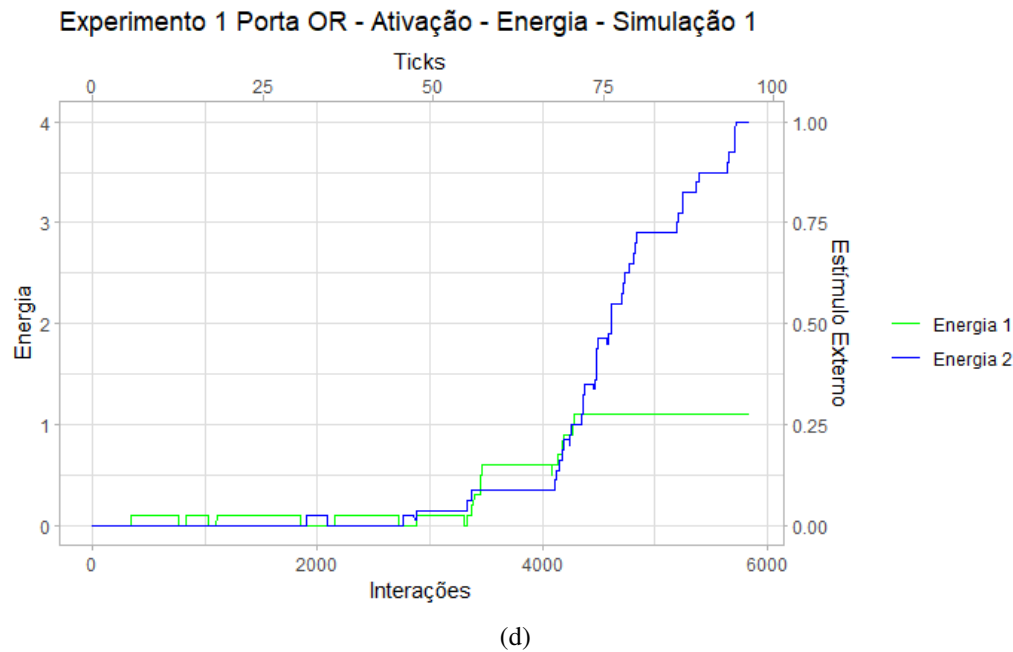
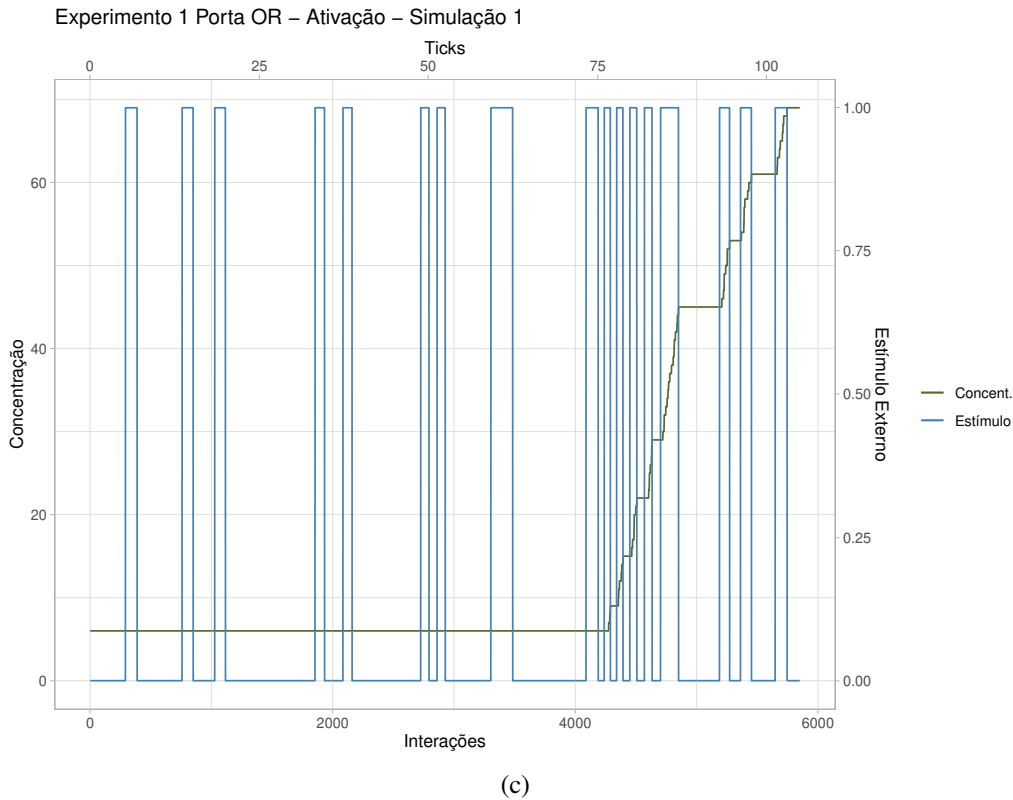


Figura 5.5: A Figura 5.5a apresenta o aumento do Produto mediante o funcionamento do Fator Externo (External Stimulus) e o funcionamento da porta AND; A Figura 5.5c apresenta o aumento do Produto de acordo com a ativação do Fator Externo e o funcionamento da Porta OR; A Figura 5.5b mostra o valor da Energia de TF1 e TF2, ou seja, o somatório das taxas de TF1 e TF2 respectivamente, de acordo com a ativação do Fator Externo e o funcionamento da porta AND; Já a Figura 5.5d mostra o valor da Energia de TF1 e TF2 de acordo com a ativação do Fator Externo e o funcionamento da porta OR.

Assim como para porta AND, foram realizados experimentos com a porta OR para função de ativação e experimentos para a função de inibição. Cada um dos experimentos possui 10 rodadas e as mesmas condições dos experimentos com a porta AND, exceto pelo fato de ser utilizada uma porta OR nas simulações.

O que difere nas configurações para simulação de inibição com a porta AND definidas na Tabela 19 é a quantidade de agentes que representam o produto na inicialização. Nesta simulação, o parâmetro “qtd” é definido como 25, porém os valores do parâmetro “qtd” para TF1 e TF2 permanecem iguais aos definidos na Tabela 15.

Tabela 19: Arquivo de entrada de definições - Inibição (porta AND) - Resultados - Comparação Uri Alon - 2

tam	cor	qtd	nome-proteína	nome-regulado	patche-cor-att	porta
1	violet	50	“X”	“Z”	orange	1
1	red	50	“Y”	“Z”	orange	1
1	brown	25	“Z”	-1	orange	1

A Tabela 20 possui as mesmas configurações da Tabela 16 exceto pelo fato de que os parâmetros “ft-1-valor” e “ft-2-valor” recebem o valor 0, indicando ação de inibição.

Tabela 20: Arquivo de entrada de restrições - Inibição (porta AND) - Resultados - Comparação Uri Alon - 2

id	ft-1	ft-1-valor	n-n1	ft-2	ft-2-valor	n-n2	regulado	patche-cor	px	py
“X”	“X”	-1	-1	“Y”	-1	-1	“Z”	-1	-1	-1
“Y”	“Y”	-1	-1	“X”	-1	-1	“Z”	-1	-1	-1
“Z”	“X”	0	1	“Y”	0	1	0	orange	0	0

O que difere nas configurações para simulação de inibição com a porta OR, definidas na Tabela 21 é a quantidade de agentes que representam o produto na inicialização. Nesta simulação, o parâmetro “qtd” é definido como 25, porém os valores do parâmetro “qtd” para TF1 e TF2 permanecem iguais aos definidos na Tabela 17.

Tabela 21: Arquivo de entrada de definições - Inibição (porta OR) - Resultados - Comparação Uri Alon - 2

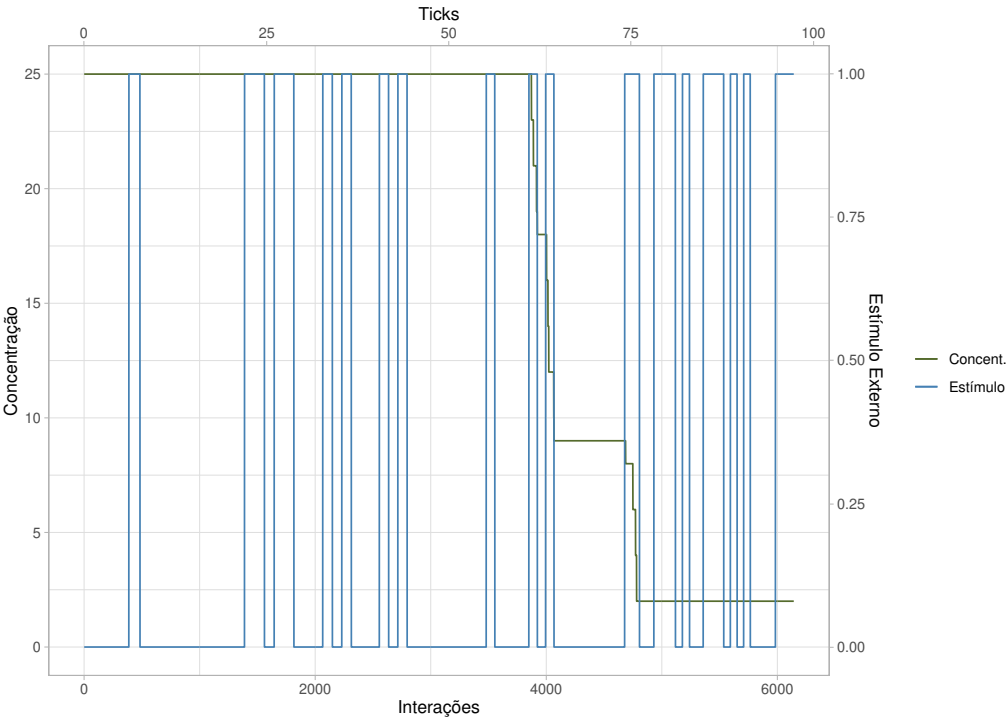
tam	cor	qtd	nome-proteína	nome-regulado	patche-cor-att	porta
1	violet	50	“X”	“Z”	orange	0
1	red	50	“Y”	“Z”	orange	0
1	brown	25	“Z”	-1	orange	0

A Tabela 22 possui as mesmas configurações da Tabela 18, exceto pelo fato de que os parâmetros “ft-1-valor” e “ft-2-valor” recebem o valor 0, indicando ação de inibição.

Tabela 22: Arquivo de entrada de restrições - Inibição (porta OR) - Resultados - Comparação Uri Alon - 2

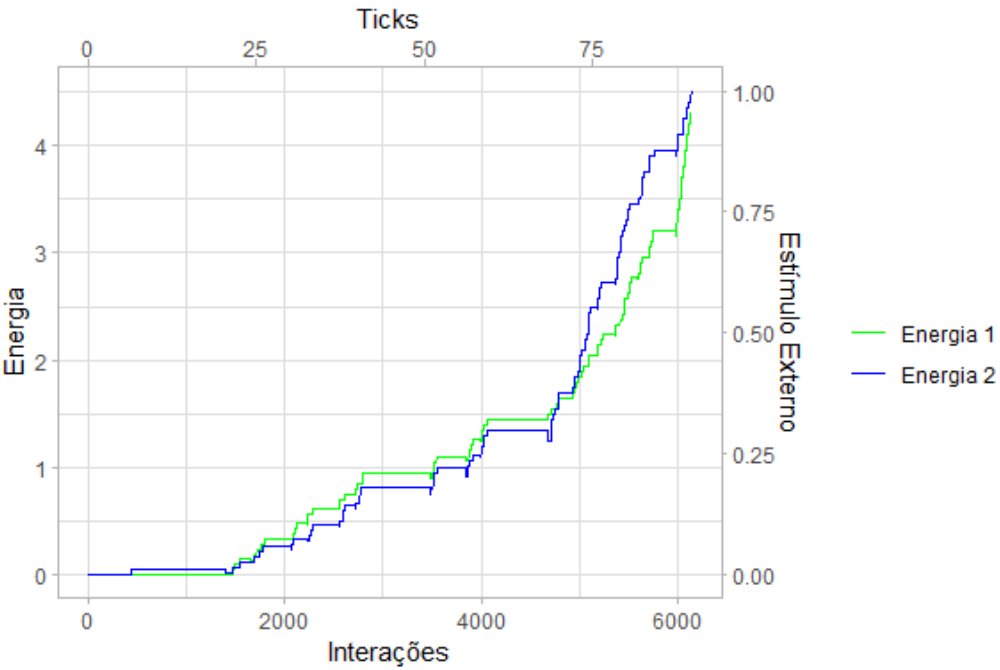
id	ft-1	ft-1-valor	n-n1	ft-2	ft-2-valor	n-n2	regulado	patche-cor	px	py
"X"	"X"	-1	-1	"Y"	-1	-1	"Z"	-1	-1	-1
"Y"	"Y"	-1	-1	"X"	-1	-1	"Z"	-1	-1	-1
"Z"	"X"	0	1	"Y"	0	1	0	orange	0	0

Experimento 1 Porta AND – Inibição – Simulação 1



(a)

Experimento 1 Porta OR - Ativação - Energia - Simulação 1



(b)

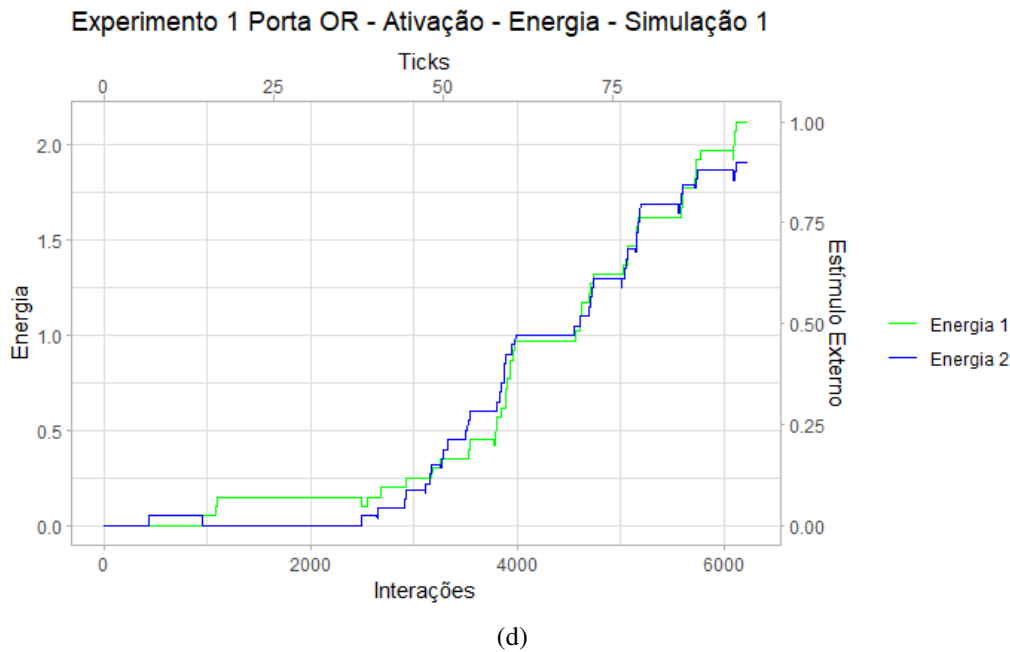
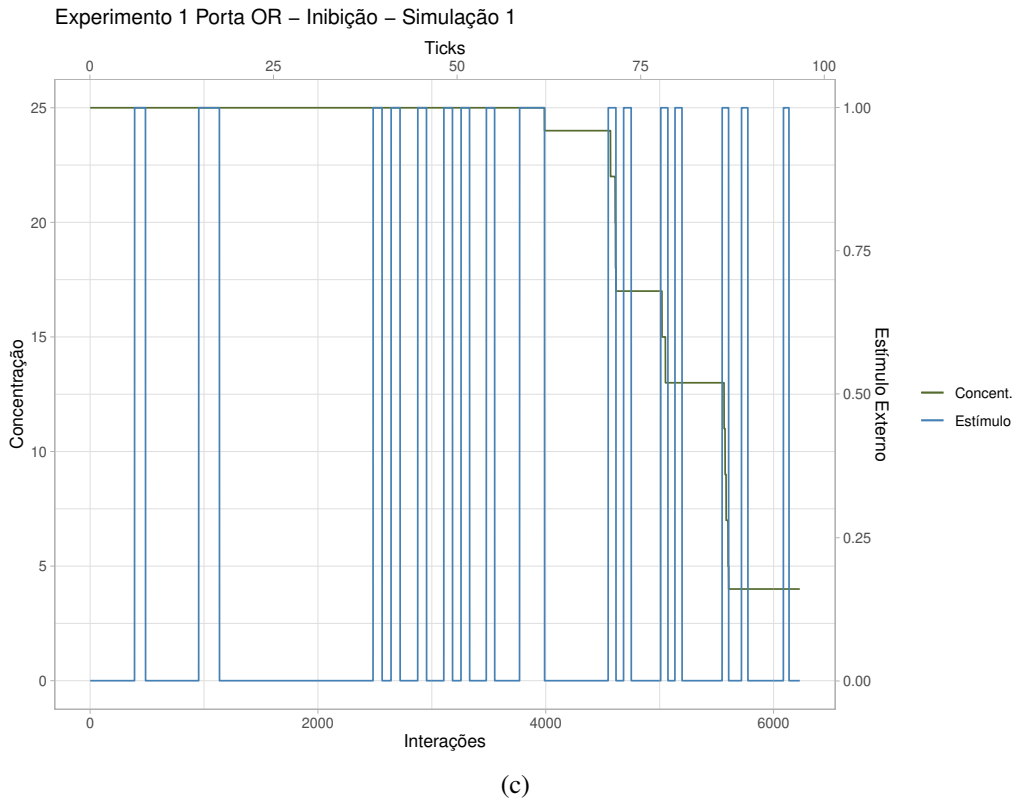


Figura 5.6: A Figura 5.6a apresenta a inibição do Produto mediante o funcionamento do Fator Externo (External Stimulus) e o funcionamento da porta AND; A Figura 5.6c apresenta o decaimento do Produto de acordo com a ativação do Fator Externo e o funcionamento da Porta OR; A Figura 5.6b mostra o valor da Energia de TF1 e TF2 de acordo com a ativação do Fator Externo e o funcionamento da porta AND; Já a Figura 5.6d mostra o valor da Energia de TF1 e TF2 de acordo com a ativação do Fator Externo e o funcionamento da porta OR.

Ao analisar a Figura 5.5 pode-se observar, primeiramente, que os gráficos encontrados nas Figuras 5.5a e 5.5c mostram que o crescimento do produto ocorre somente em passos em que o fator externo

está ativado, observando-se as Figuras 5.5a e 5.5b pode-se constatar que o crescimento em simulações com a porta AND depende de o fator externo estar ativado e também de haver a quantidade mínima de energia de TF1 e TF2 no sítio de ligação para que ocorra a função bioquímica que resulta no aumento desse Produto (tam3).

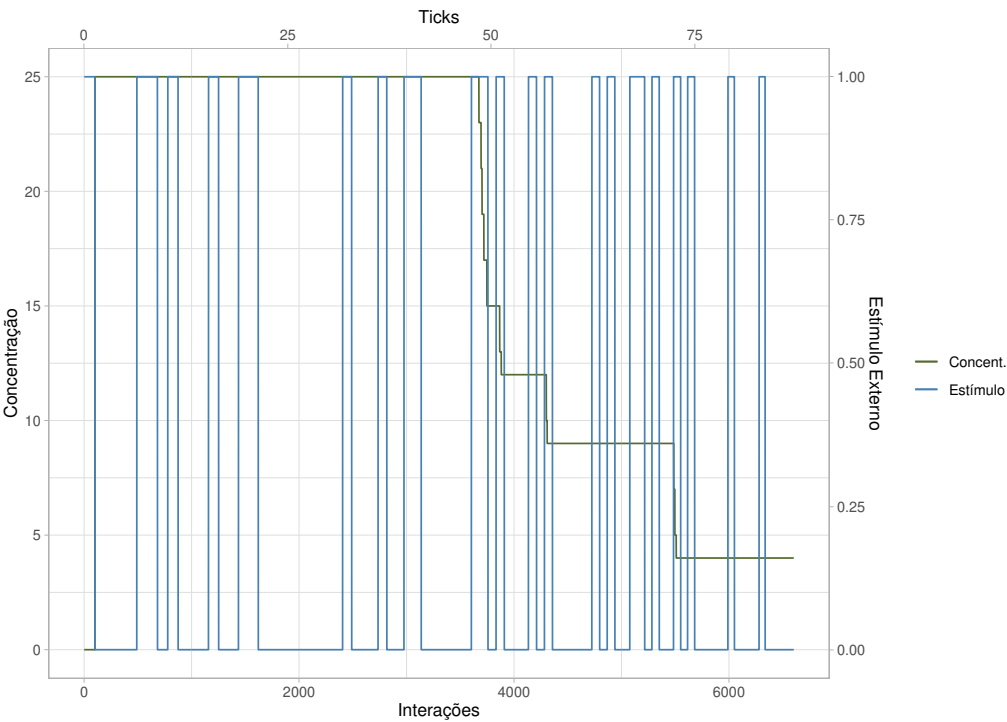
Também observa-se nas Figuras 5.5c e 5.5d que o crescimento do produto em simulações com a porta OR ocorre quando o fator externo está ativo e quando ao menos uma das energias dos fatores de transcrição no sítio de ligação tem o valor mínimo estabelecido para desencadear a função de produção.

Continuando a análise da Figura 5.5, pode-se comparar as Figuras 5.5a e 5.5c, e percebe-se que o crescimento do produto ocorre muito mais rápido na simulação com a porta OR, como era esperado.

Ao analisar a Figura 5.6 pode-se observar que os gráficos encontrados nas Figuras 5.6a e 5.6c mostram que o decrescimento do produto ocorre somente em passos em que o fator externo está ativado, observando-se as Figuras 5.6a e 5.6b pode-se perceber que a inibição em simulações com a porta OR também dependem de o fator externo estar ativado, mas de somente haver a quantidade mínima de energia de um dos fatores de transcrição (TF1 ou TF2) no sítio de ligação para que ocorra a função bioquímica, que resulta no aumento desse produto (tam3). O mesmo comportamento ocorre para simulações com a porta OR, como pode ser visto nas Figuras 5.6c e 5.6d.

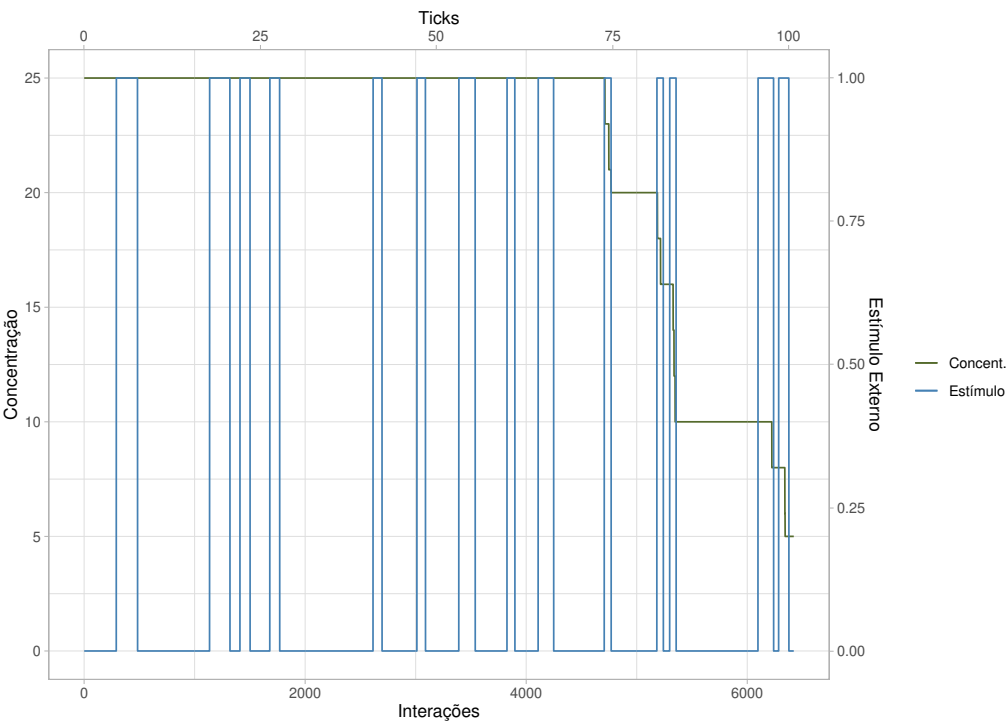
Analisando a Figura 5.6, pode-se comparar as Figuras 5.6a e 5.6c e percebe-se que o crescimento do produto não ocorre muito mais rápido na simulação com a porta OR nesta rodada deste experimento como era esperado, porém pode-se observar o comportamento esperado em outras rodadas deste mesmo experimento e em rodadas dos outros dois experimentos realizados com a porta OR. Já na Figura 5.7 pode-se visualizar à esquerda os gráficos do Fator Externo versus Produto de simulações com a porta AND e à direita os gráficos com simulações de porta OR, pode-se constatar que nestas simulações o decaimento ocorre mais rápido em simulações com a porta OR de acordo com o comportamento esperado.

Experimento 1 Porta AND – Inibição – Simulação 2



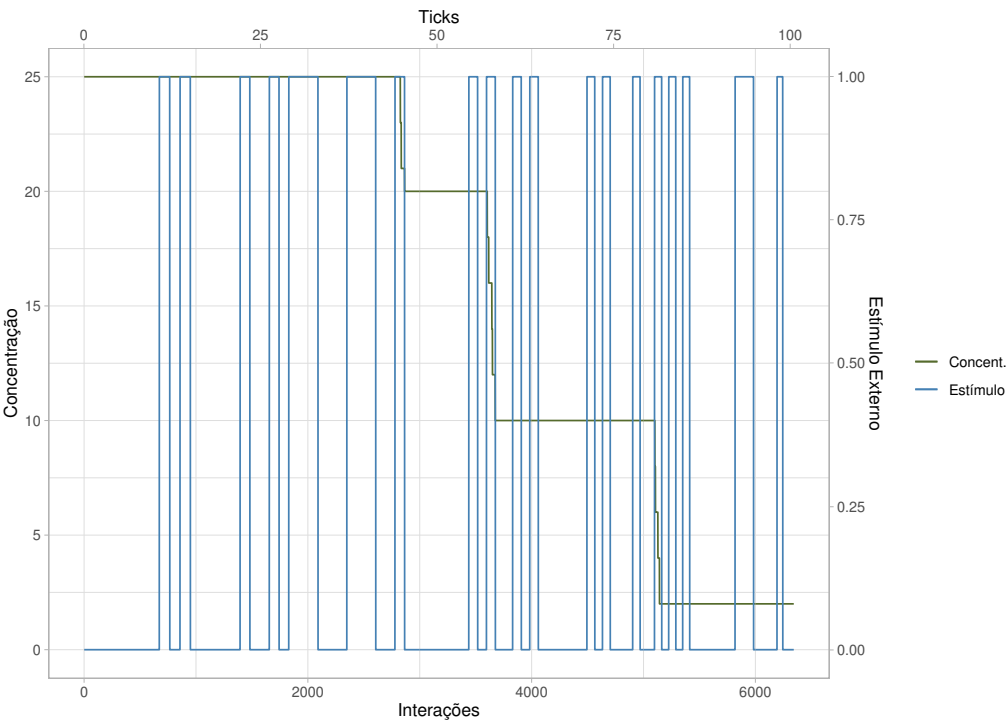
(a)

Experimento 1 Porta OR – Inibição – Simulação 2



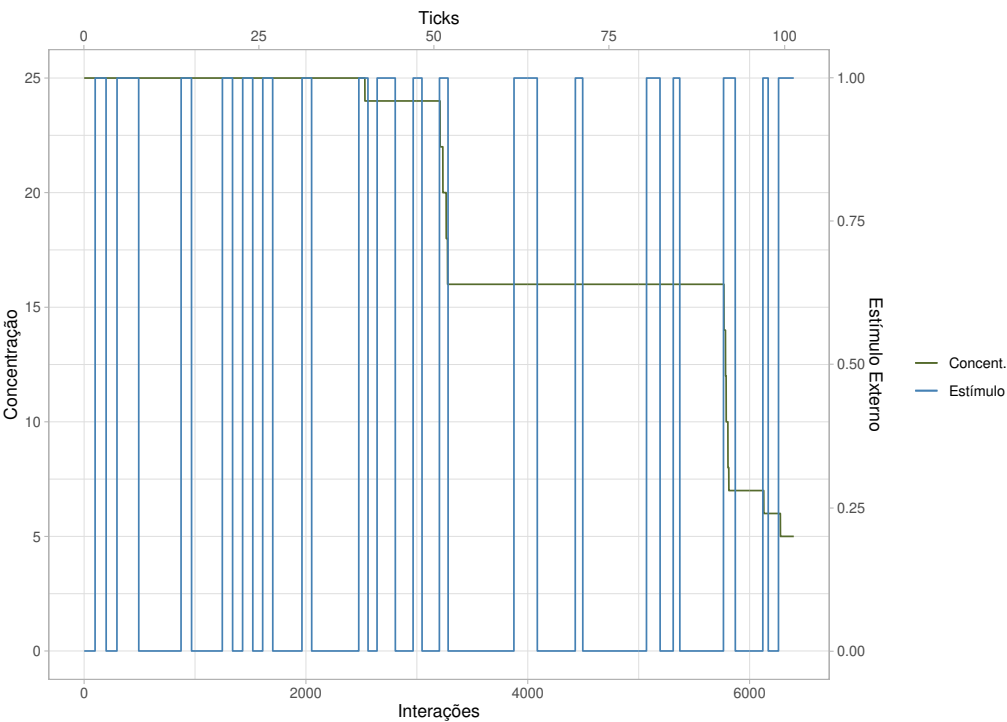
(b)

Experimento 1 Porta AND – Inibição – Simulação 3



(c)

Experimento 1 Porta OR – Inibição – Simulação 3



(d)

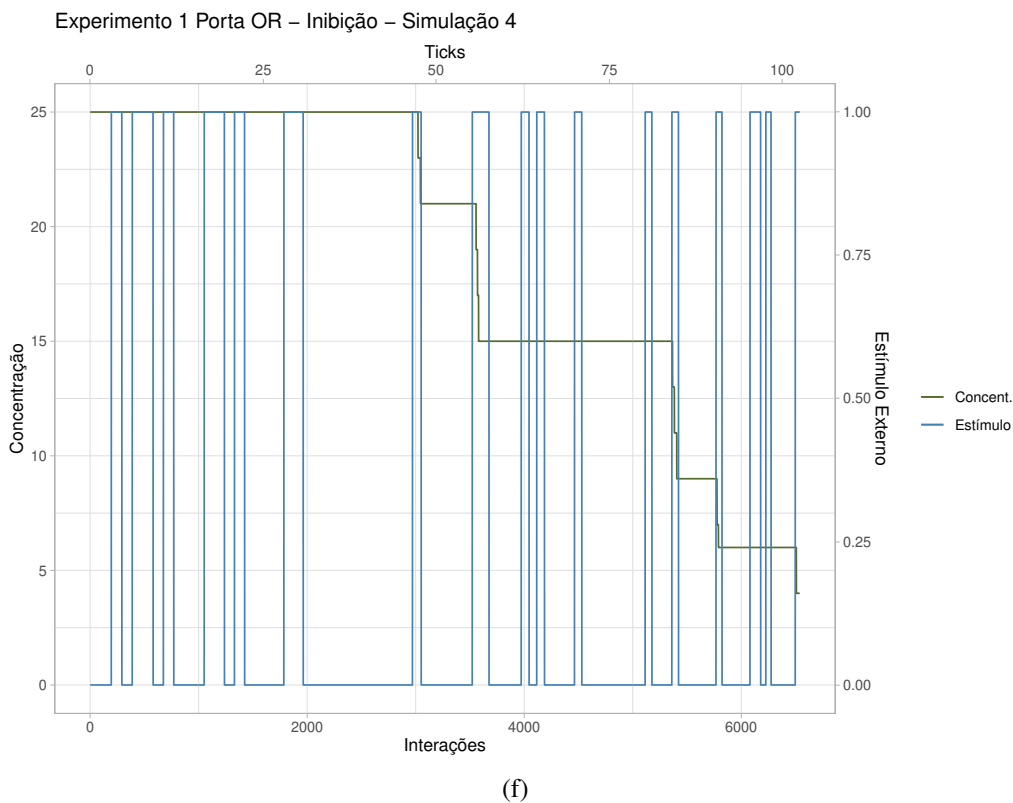
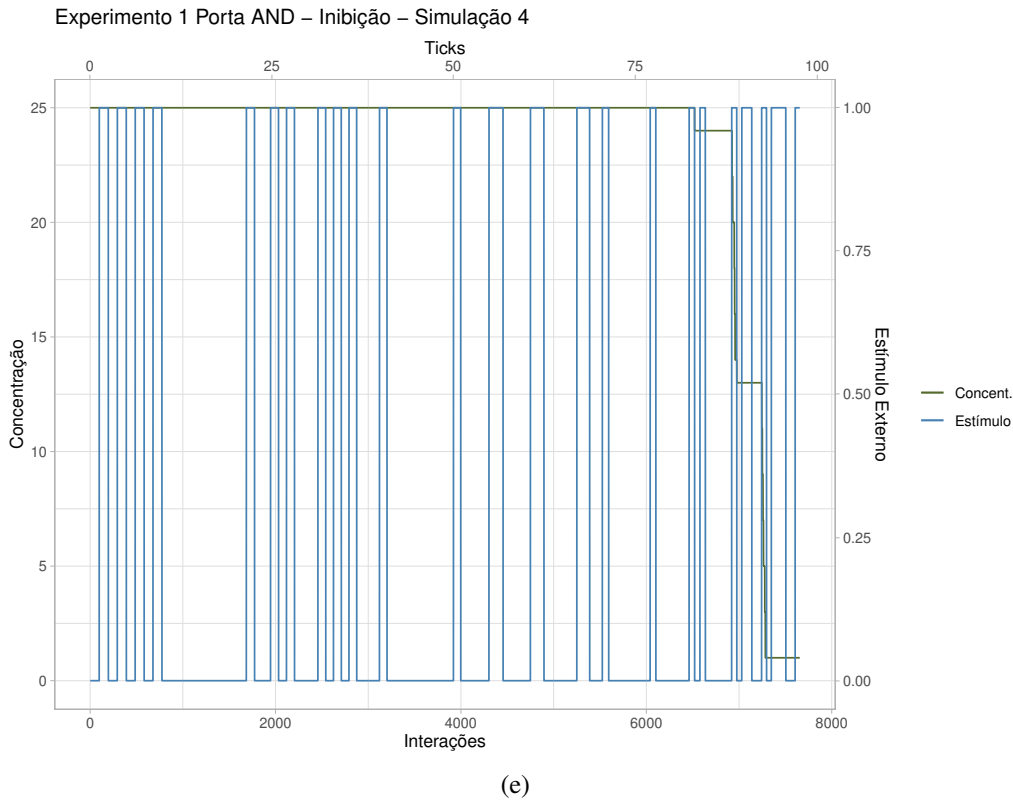


Figura 5.7: As Figuras 5.7a, 5.7b, 5.7c, 5.7d, 5.7e, 5.7f apresenta os gráficos do Fator Externo das rodadas 2,3,4 e 5 do experimento 1 com a porta AND e das rodadas 2,3,4 e 5 do experimento 2 com a porta OR.

A Figura 5.8 mostra o funcionamento do experimento de Uri Alon. E pode-se observar que quando o sinal S_x aparece, X se torna ativo e rapidamente se liga a seus promotores downstream.

Como resultado, Y começa a se acumular. No entanto, devido à função de entrada AND, a produção de Z começa apenas quando a concentração de Y cruza o limite de ativação para o promotor Z. Isso resulta em um retardo da expressão de Z após o aparecimento de S_x [Alon, 2007].

Ao analisar a Figura 5.8 e comparar com as Figuras 5.5 e 5.6, pode-se verificar que o gráfico do modelo de Uri Alon [Alon, 2007] e os gráficos resultantes deste modelo não possuem a mesma apresentação, porém os gráficos deste modelo encontrados em 5.5 e 5.6 mostram que o Produto somente é produzido ou inibido quando o Fator Externo está ativo e quando possui energia necessária suficiente no sítio de ligação, assim como o funcionamento do modelo de Uri Alon.

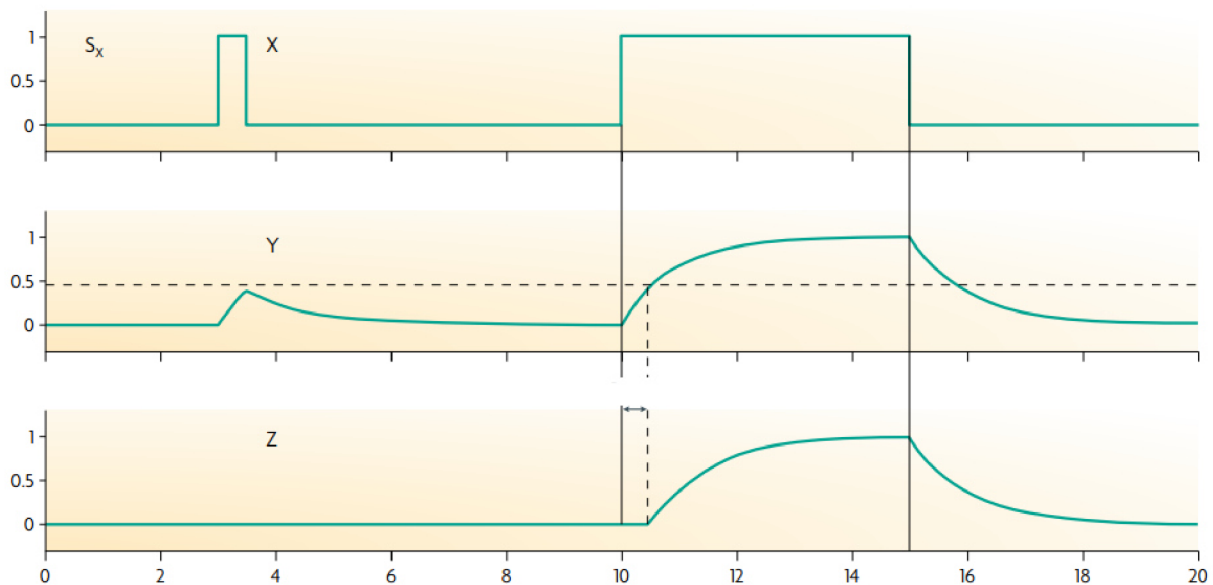


Figura 5.8: Como pode ser observado neste gráfico, este motif com uma função de entrada AND mostra atraso após a adição do estímulo (S_x) e nenhum atraso após a remoção do estímulo. Sendo X e Y fatores de transcrição e Z o produto. Portanto, atua como um filtro sensível ao sinal, que responde apenas a estímulos persistentes [Alon, 2007].

O modelo de Uri Alon [Alon, 2007] considera os valores das taxas dos Fatores de Transcrição dentro do sítio de ligação e essas taxas crescem quando o Fator Externo está ativado e decrescem quando o Fator Externo encontra-se desativado. Neste contexto, não é considerado que pode haver mais de uma unidade de um mesmo Fator de Transcrição no sistema, também não é considerado que os Fatores de Transcrição podem entrar no sítio de ligação de uma forma mais aleatória durante o decorrer do processo [Alon, 2007].

Na Figura ?? pode-se observar o comportamento das taxas dos Fatores de Transcrição porém, percebe-se que há vezes em que o Fator Externo está ativado, mas não há registro de atividade de nenhum dos Fatores de Transcrição e por outras vezes não decrescimento da energia. Tal comportamento ocorre devido ao fato de o movimento dos Fatores de Transcrição no ambiente (grid da simulação) ser aleatório. Sendo Assim, há momentos em que o Fator Externo está ativado, mas não há Fatores de Transcrição no sítio de ligação e também por vezes pode ocorrer de sempre estar entrando um Fator de Transcrição no sítio de ligação.

5.3 Comparação com o Modelo da planta *Arabidopsis Thaliana*

5.3.1 Metodologia

O Modelo Proposto para realização de simulações de GRN foi desenvolvido baseado na versão anterior que foi baseada no Modelo de Uri Alon. Esta versão foi desenvolvida baseada em “motifs”.

É possível configurar as interações da rede através do modelo de “motifs” em arquivos de entrada, além disso é possível alterar as taxas de crescimento e decaimento de energia diretamente na interface do modelo. Este modelo possui flexibilidade para configurar vários tipos de interações possíveis e vários tipos de redes GRN. A flexibilidade para configuração das simulações facilita a realização de simulações de GRN genéricas, sendo os arquivos de entrada facilmente alteráveis.

O modelo possui dois arquivos de entrada baseado no padrão de “motifs” para que seja possível realizar a configuração das interações que podem ocorrer na simulação.

A versão apresentada na Figura ?? mostra as configurações iniciais para a simulação do ciclo da planta *Arabidopsis Thaliana*.

Mediante uma análise dos esquemas apresentados nas Figuras 2.9 e 2.10 foi realizada a configuração dos arquivos de entrada necessários para efetuar a simulação da GRN que representa o ciclo da planta *Arabidopsis Thaliana*. Através da análise do modelo da planta *Arabidopsis Thaliana* chegou-se na configuração de “motifs” para realizar os ajustes e calibração para realização das simulações apresentada na Tabela 24.

Tabela 23: Genes/Fatores de transcrição do ciclo da *Arabidopsis Thaliana*. Apresenta os fatores de transcrição e suas respectivas cores na interface do netlogo

Cor	Fator de Transcrição
”violet”	”LHY/CCA1”
”magenta”	”LHYmod”
”red”	”PRR9”
”brown”	”PRR7”
”pink”	”NI”
”orange”	”GI”
”gray”	”Y”
”yellow”	”TOC1”
”cyan”	”ZTL”
”sky”	”TOC1mod”

A Tabela 23 apresenta os Fatores de Transcrição envolvidos no ciclo da planta *Arabidopsis Thaliana* e suas respectivas cores de representação neste modelo.

Tabela 24: Interação - Tradução para “motifs”

Interação	Tradução para “Motif”	Sítio de Ligação
LHY/CCA1 → LHYmod	LHY/CCA1 OR LHY/CCA1 → LHYmod	yellow
LHY/CCA1 → PRR9	LHY/CCA1 OR LHY/CCA1 → PRR9	brown
LHY/CCA1 → PRR7	LHY/CCA1 OR LHY/CCA1 → PRR7	orange
LHYmod → NI	LHYmod OR LHYmod → NI	sky
LHY/CCA1 ⇝ GI	LHY/CCA1 OR LHY/CCA1 ⇝ GI	blue
LHY/CCA1 ⇝ Y	LHY/CCA1 OR LHY/CCA1 ⇝ Y	red
LHY/CCA1 ⇝ TOC1	LHY/CCA1 OR LHY/CCA1 ⇝ TOC1	gray
TOC1 ⇝ GI	TOC1 OR TOC1 ⇝ GI	pink
TOC1 ⇝ Y	TOC1 OR TOC1 ⇝ Y	orange
Y → TOC1	Y OR Y → TOC1	37
Y → GI	Y OR Y → GI	27
GI → ZTL	GI OR GI → ZTL	violet
ZTL ⇝ TOC1	ZTL OR ZTL ⇝ TOC1	white
TOC1 ⇝ PRR9	TOC1 OR TOC1 ⇝ PRR9	99
PRR9 ⇝ LHY/CCA1	PRR9 OR PRR9 ⇝ LHY/CCA1	19
PRR7 ⇝ LHY/CCA1	PRR7 OR PRR7 ⇝ LHY/CCA1	63
NI ⇝ LHY/CCA1	NI OR NI ⇝ LHY/CCA1	turquoise
TOC1 → TOC1mod	TOC1 OR TOC1 → TOC1mod	lime
TOC1mod → LHY/CCA1	TOC1mod OR TOC1mod → LHY/CCA1	magenta
GI → TOC1	GI OR GI → TOC1	47

A Tabela 24 apresenta as relações existentes no ciclo da Arabidopsis Thaliana em forma de “motifs”. Essas relações foram então traduzidas para operações com a porta OR. Então, esta tabela possui as interações a tradução para forma de “motifs” e sítio de ligação onde a interação ocorre.

Tabela 25: Recorte de um arquivo de entrada - Atributos

Tam	Cor	Quantidade	FT	Produto	Cor SL	Porta	Luz	Período
1	violet	10	“LHY-yellow”	“LHYmod-yellow”	yellow	0	1	1
1	violet	10	“LHY-yellow”	“LHYmod-yellow”	yellow	0	1	1
1	magenta	2	“LHYmod-yellow”	-1	yellow	0	1	1
1	violet	10	“LHY-brown”	“PRR9-brown”	brown	0	1	1
1	violet	10	“LHY-brown”	“PRR9-brown”	brown	0	1	1
1	red	2	“PRR9-brown”	-1	brown	0	1	1
1	violet	10	“LHY-orange”	“PRR7-orange”	orange	0	1	1
1	violet	10	“LHY-orange”	“PRR7-orange”	orange	0	1	1
1	brown	2	“PRR7-orange”	-1	orange	0	1	1

A Tabela 25 apresenta um exemplo do arquivo de entrada utilizado para configurar os atributos do modelo (o arquivo completo, bem como o código do framework, arquivos de saída e gráficos encontram-se no link https://github.com/Nilzair/Tese_NilzairAgostinho_-Framework_GRN_Files). Esta tabela contém o tamanho do agente/gene/fator de transcrição na interface gráfica do Netlogo, a cor do agente, quantidade inicial, o nome do fator de transcrição, cor do Sítio de Ligação (região quadrada na interface do netlogo onde ocorrem as interações no modelo), a porta de interação

(OR/AND), se o fator de transcrição recebe incidência de luz e em que período que recebe a incidência de luz.

Tabela 26: Recorte de um arquivo de entrada - Restrições

ID	FT 1	A/I FT 1	Qtd de FT 1	FT 2	A/I FT 2	Qtd de FT 2	Produto	Cor SL	dx	dy
“LHY-yellow”	“LHY-yellow”	-1	-1	“LHY-yellow”	-1	-1	“LHYmod”	-1	-1	-1
“LHY-yellow”	“LHY-yellow”	-1	-1	“LHY-yellow”	-1	-1	“LHYmod”	-1	-1	-1
“LHYmod-yellow”	“LHY-yellow”	1	1	“LHYmod-yellow”	1	1	-1	yellow	-22	20
“LHY-brown”	“LHY-brown”	-1	-1	“LHY-brown”	-1	-1	“PRR9”	-1	-1	-1
“LHY-brown”	“LHY-brown”	-1	-1	“LHY-brown”	-1	-1	“PRR9”	-1	-1	-1
“PRR9-brown”	“LHY-brown”	1	1	“PRR9-brown”	1	1	-1	brown	-8	20
“LHY-orange”	“LHY-orange”	-1	-1	“LHY-orange”	-1	-1	“PRR7”	-1	-1	-1
“LHY-orange”	“LHY-orange”	-1	-1	“LHY-orange”	-1	-1	“PRR7”	-1	-1	-1
“PRR7-orange”	“LHY-orange”	1	1	“PRR7-orange”	1	1	-1	orange	8	20

A Tabela 26 possui as informações necessárias das restrições aplicadas para calibração do modelo. A primeira coluna possui o que foi atribuído como ID do Fator de Transcrição, a seguir o nome do Fator de Transcrição 1 e se é ação de ativação ou inibição que é seguido da quantidade necessária do FT para que aja interação. E após as mesmas informações para o FT 2 seguido do produto da interação, cor do sítio de ligação e posicionamento do sítio de ligação (patche) na interface.

5.3.2 Resultados

Este estudo de caso é validado através da comparação com o Modelo da planta *Arabidopsis Thaliana* [Hearn et al., 2018], porém, como dito anteriormente, o framework foi totalmente baseado na teoria dos “motifs” [Alon, 2007].

Foram configurados 20 “motifs”, ou seja 20 sítios de ligação nos arquivos de entrada. Cada três linhas completam o conjunto de configurações de um “motif”, que envolve dois TFs e um produto. A porta utilizada para configurar esse ciclo foi a porta OR e existem ações de ativação e inibição que dependem do período em que se encontram, do período em que são acionadas configurado no arquivo de entrada e se necessitam ou não de luz para que ocorra a ação.

Em cada linha que representa um gene/produto nos arquivos de entrada foram configurados para iniciar a simulação com **22 genes/agentes** que representam 3% da área do grid (750 - 30 x 25). Cada sítio de ligação foi configurado com uma cor diferente o tamanho (3 x 3).

Para cada ação é necessário ter apenas um gene/agente de cada TF no respectivo sítio de ligação, mas é preciso ter no mínimo o produto ativo de 1. Porém, todos os agentes tem seu “produto ativo” decrementado de acordo com o valor configurado na interface do framework durante todo o tempo da simulação, o que simula a perda de energia que ocorre na natureza.

Em um dos arquivos de entrada é possível configurar se ação que ocorre em cada “motif” depende da incidência de luz e em que período pode ocorrer. Então, no arquivo de entrada são definidos “period” 1 para “morning”, “period” 2 para “central”, “period” 3 para “vespertine”, “period” 4 para “nocturnal” e “period” 5 quando a ação pode ocorrer em qualquer período. Assim, pode-se verificar de acordo com a Tabela 24 que possui a definição das interações do ciclo que alguns genes tendem a somente decrescer o produto ativo, devido a definições de exposição à luz, período em que ocorre a ação, questão espacial como tamanho do grid, tamanho do Sítio de Ligação e movimentação dos agentes de forma aleatória.

Devido a localização espacial, taxa de decaimento e a quantidade de agentes pode-se ter um ambiente mais “denso” o que pode culminar com mais ativações e/ou inibições. Sendo assim, o framework proporciona a definição de modelos em ambientes muito flexíveis.

Na Figura ?? pode-se verificar média dos resultados do Modelo da Planta *Arabidopsis Thaliana* e percebe-se que a variação do “produto ativo” é pequena, assim como nos resultados obtidos desse estudo caso simulado pelo framework desenvolvido neste trabalho. Nesta mesma figura percebe-se um comportamento de oscilação, assim como é visível nos resultados apresentados a seguir, com exceção de alguns que tendem a ser inibidos devido as configurações espaciais, incidência de luz e participação da ação em mais de um sítio de ligação.

As figuras a seguir apresentam os gráficos dos comportamentos dos “produtos ativo” dos genes do ciclo circadiano da Planta *Arabidopsis Thaliana*, considerando as configurações acima citadas e as configurações nos arquivos de entrada decorrentes das configurações que encontram-se na Tabela 24. A simulação foi executada até o 300 ticks, ou seja, representando 3 dias, de acordo com a “escala” que mostra o comportamento dos períodos do dia, conforme explicado na Figura 4.4. Como dito anteriormente, o decaimento ocorre durante toda a simulação, independentemente da localização espacial ou configurações.

As escalas do eixo y (Concentração) de todos os gráficos começam da quantidade inicial do “produto ativo” de cada gene, para deixar mais evidente o comportamento de oscilação e os períodos em que podem ocorrer as ações, de acordo com as configurações de cada gene e sítio de ligação.

A Figura 5.9 representa o gráfico do gene LHYmod, que é ativado somente sob as condições de incidência de luz, energia de pelo menos 1 do gene LHY/CCA1 no sítio de ligação referente e encontrar-se no período matutino (“morning” - até tick 25). Sendo assim, o comportamento mostra-se como o esperado e semelhante ao comportamento que ocorre na natureza.

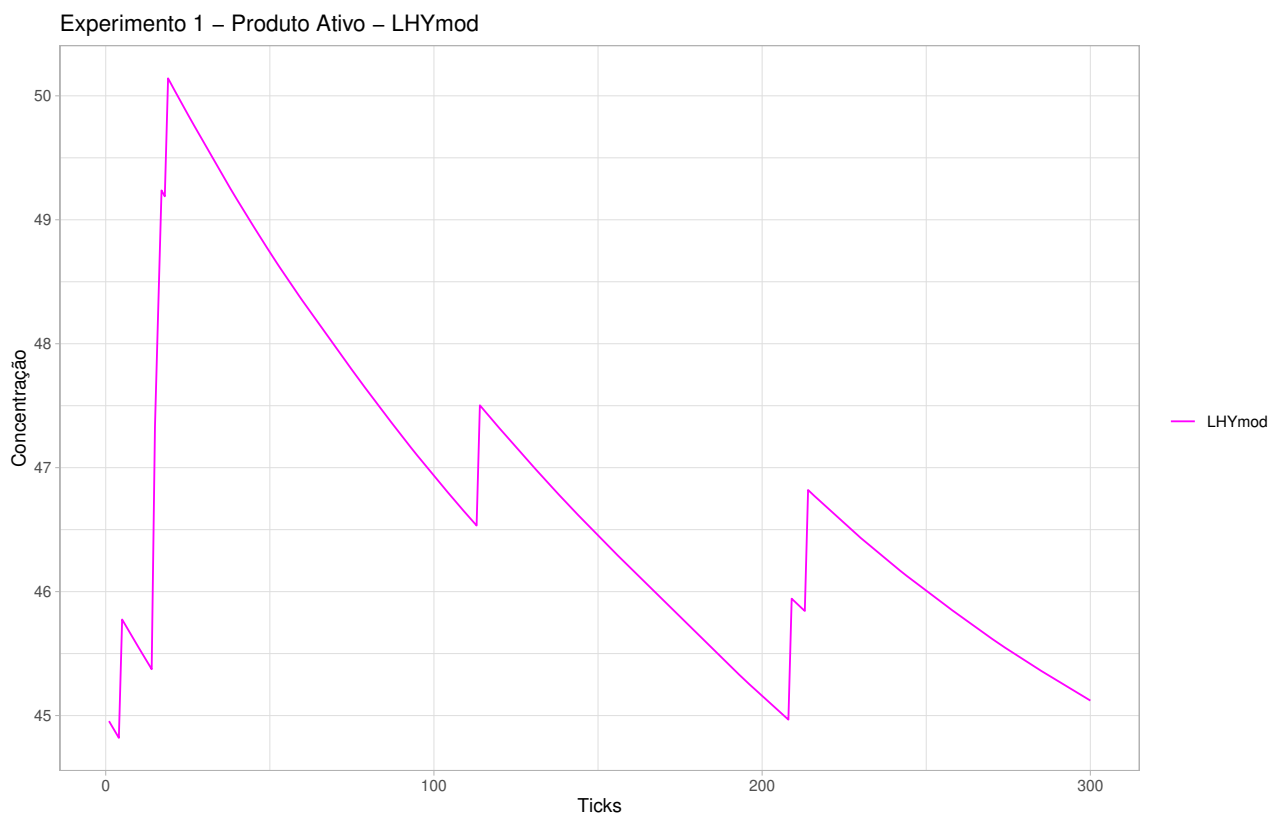


Figura 5.9: Concentração do gene LHYmod produzida

A Figura 5.10 representa o gene PRR9, que assim como o LHYmod, é ativado somente sob as condições de incidência de luz, energia de pelo menos 1 do gene LHY/CCA1 no sítio de ligação referente e encontrar-se no período matutino (“morning” - até tick 25), mas também sofre ação de inibição pelo gene TOC1 no período noturno (“nocturnal” - dos ticks 75 a 100 ticks), o que gera os “picotados” no gráfico próximo ao final do período noturno.

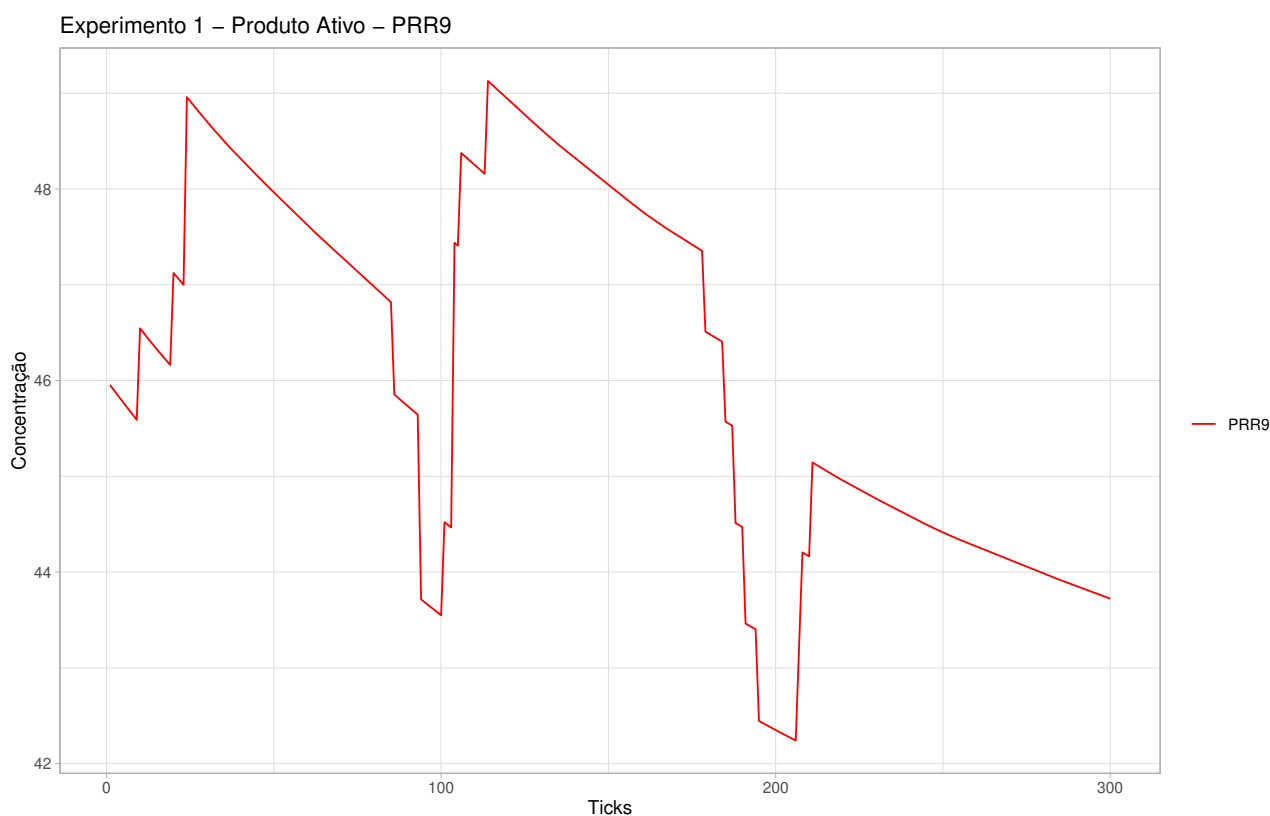


Figura 5.10: Concentração do gene PRR9 produzida

A Figura 5.11 representa o gene PRR7, que assim como os anteriores, é ativado somente sob as condições de incidência de luz, energia de pelo menos 1 do gene LHY/CCA1 no sítio de ligação referente e encontrar-se no período matutino (“morning” - até tick 25).

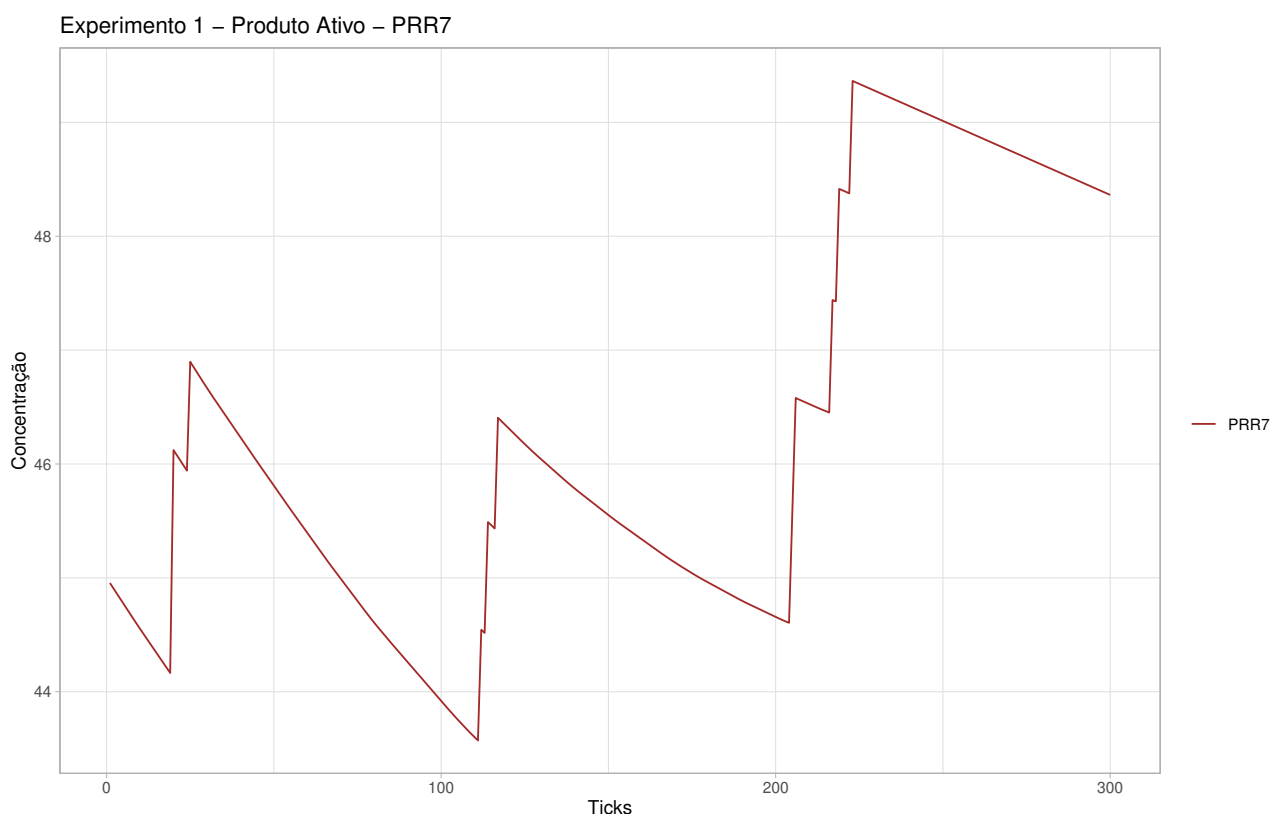


Figura 5.11: Concentração do gene PRR7 produzida

A Figura 5.12 representa o gene NI, que assim como os anteriores, é ativado somente sob as condições de incidência de luz, energia de pelo menos 1 do gene LHY/CCA1 no sítio de ligação referente e encontrar-se no período matutino (“morning” - até tick 25), e apresenta um comportamento semelhante e possui as mesmas configurações. Contudo, existe diferenças devido à localização espacial e à movimentação aleatória dos genes.

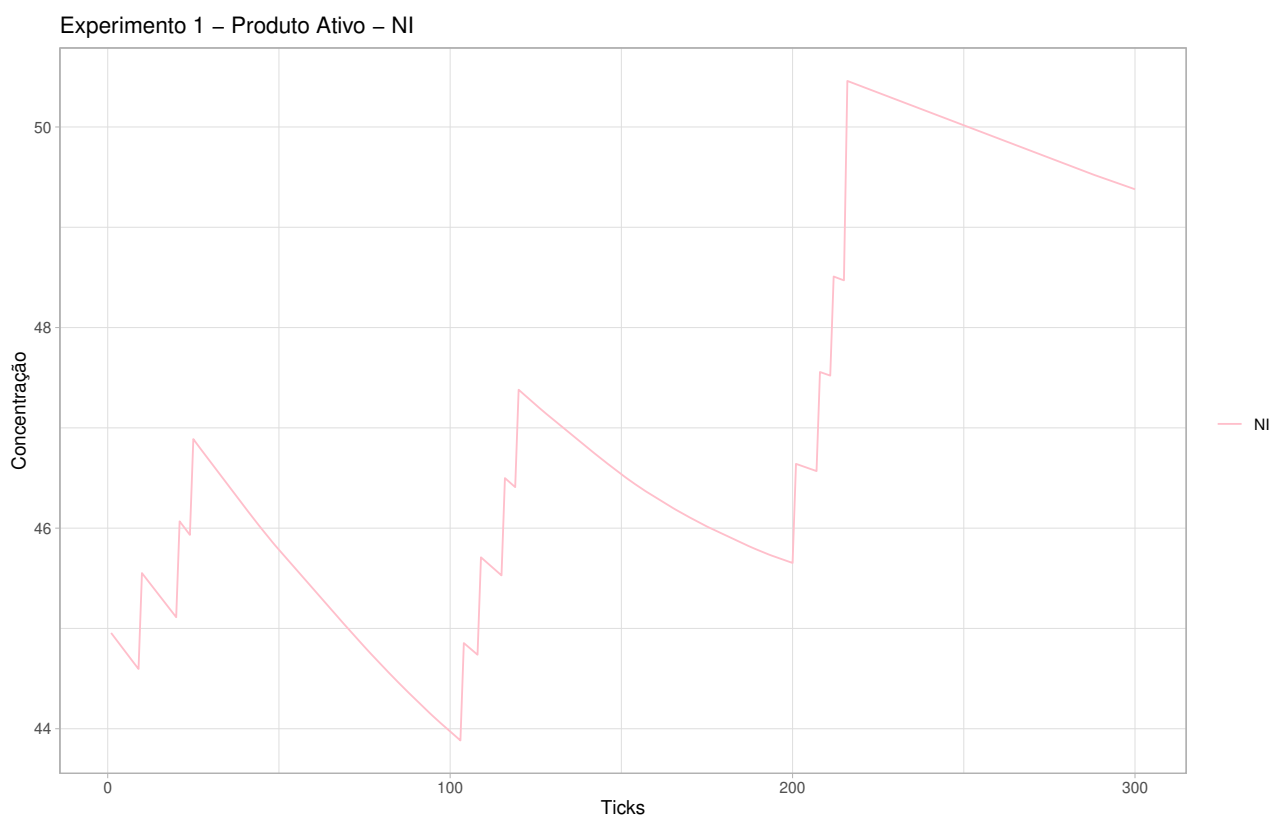


Figura 5.12: Concentração do gene NI produzida

A Figura 5.13 representa o gene GI, que é inibido sob as condições de incidência de luz, energia de pelo menos 1 do gene LHY/CCA1 no sítio de ligação referente e encontrar-se no período matutino (“morning” - até tick 25), no período noturno é inibido sob as condições de energia de pelo menos 1 do gene TOC1 no sítio de ligação referente e encontrar-se em (“nocturnal” - ticks 75 a 100), mas também é ativado pelo gene Y no período noturno com incidência remanescente da luz. Portanto, este gráfico apresenta maior comportamento de decaimento e com alguns “picos” de ativação em período noturno.

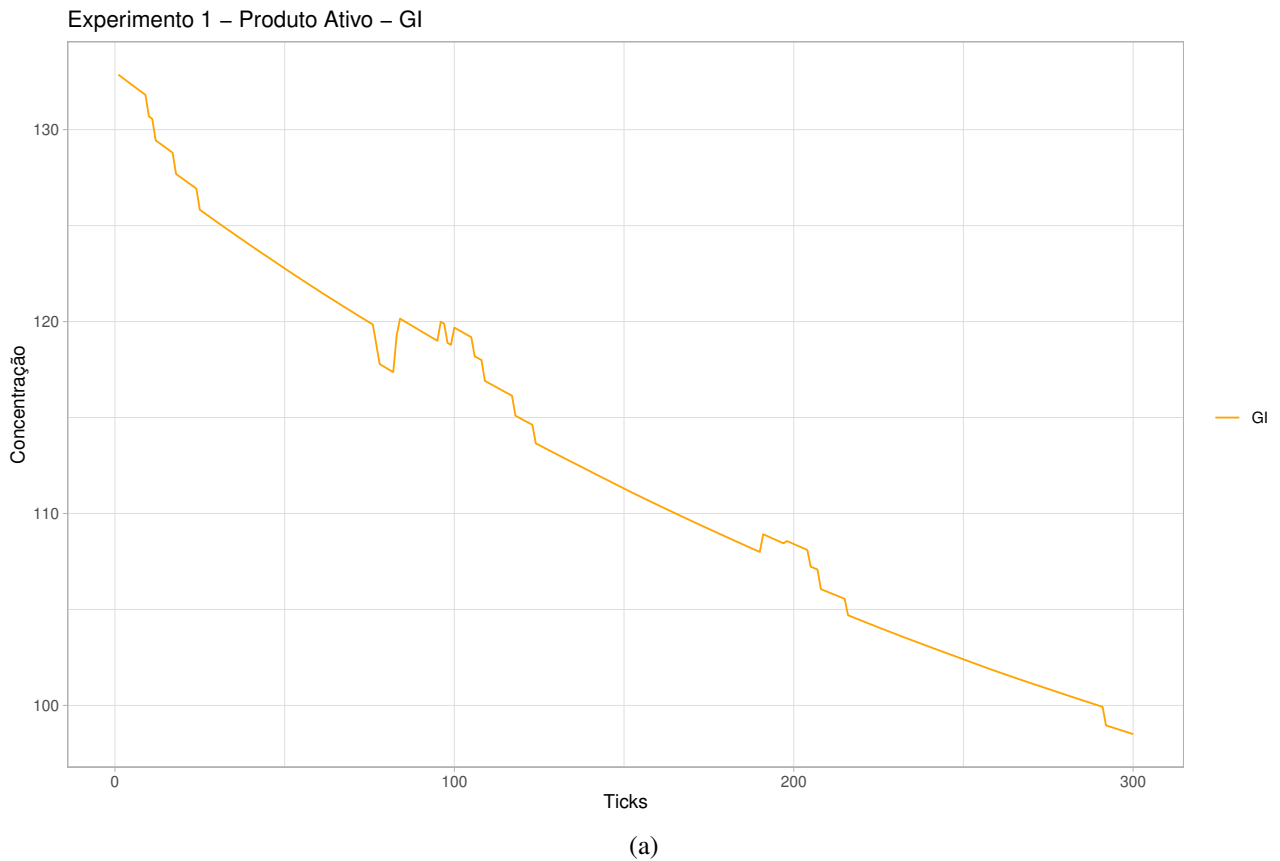


Figura 5.13: Concentração do gene GI produzida

A Figura 5.14 representa o gene TOC1mod é ativado somente sem incidência de luz, com energia de pelo menos 1 do gene TOC1 no sítio de ligação referente e encontrar-se no período noturno (“noturnal” - ticks 75 a 100). Sendo assim, apresentando decaimento nos demais períodos e ativação no período noturno.

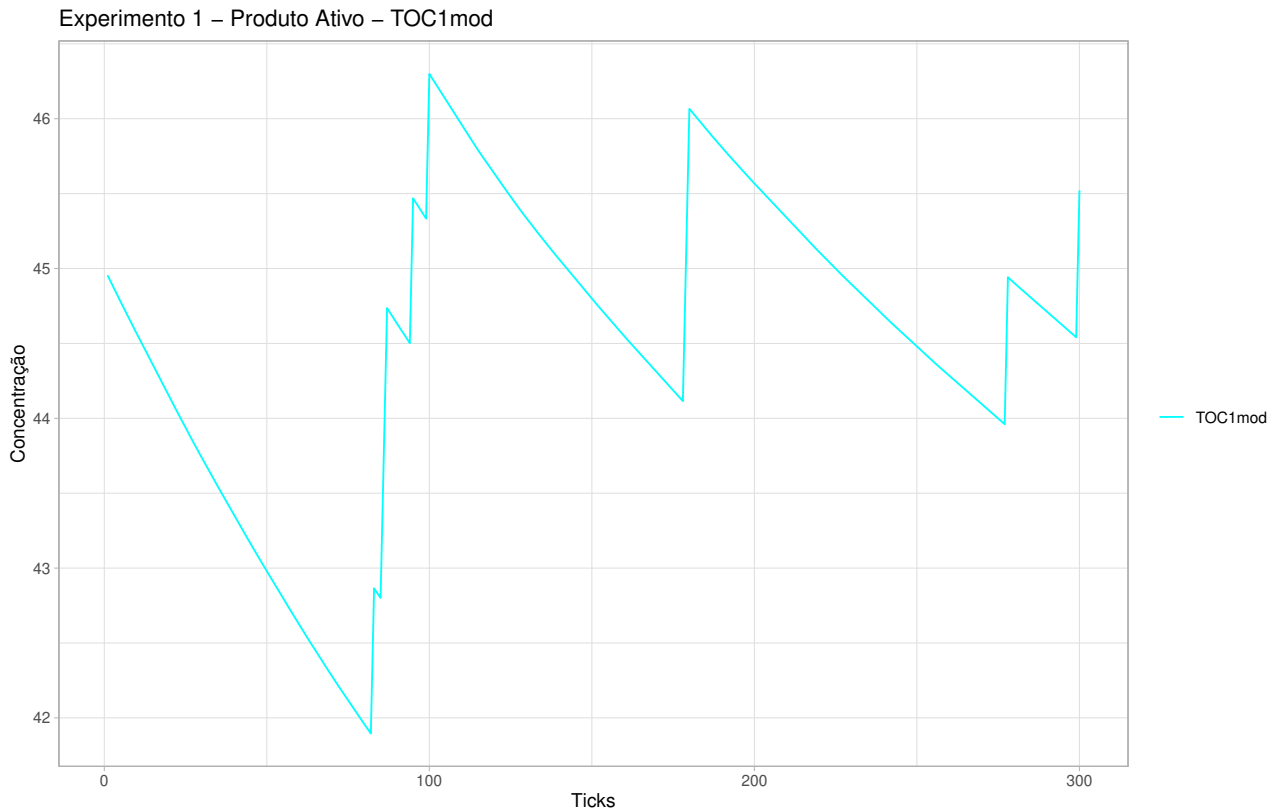


Figura 5.14: Concentração do gene TOC1mod produzida

A Figura 5.15 mostra o gráfico referente ao gene Y, que é inibido pelo gene LHY/CCA1 no período matutino (“morning” - até tick 25) sob incidência de luz e pelo TOC1 no período vespertino (“vespertine” - ticks 50 a 75), sem incidência de luz, portanto, mostrando somente decaimento.

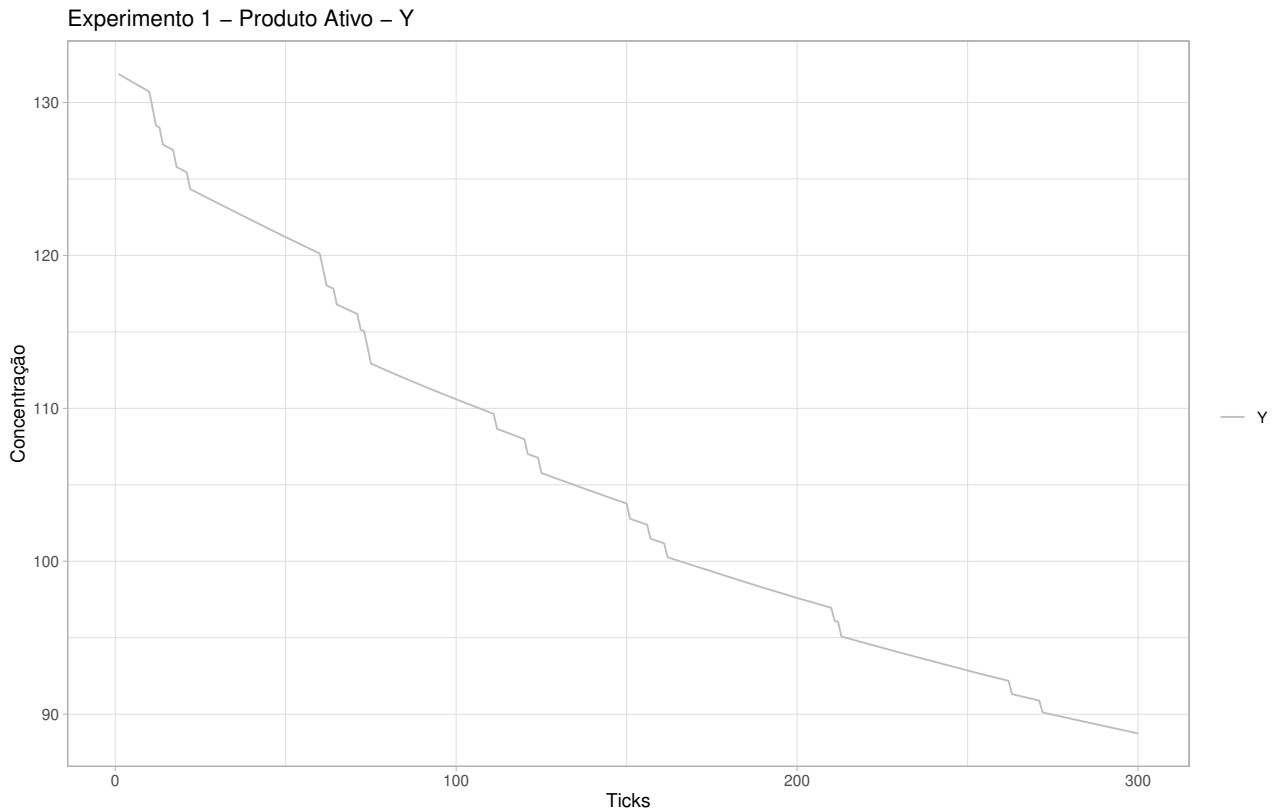


Figura 5.15: Concentração do gene Y produzida

A Figura 5.16 mostra o gráfico do gene TOC1, que é inibido pelo LHY/CCA1 no período vespertino (“vespertine” - ticks 50 a 75) sob incidência de luz, inibido pelo ZTL no período noturno (“nocturnal” - ticks 75 a 100) e ativado pelo GI no período noturno (“nocturnal” - ticks 75 a 100), sob remanescente de luz.

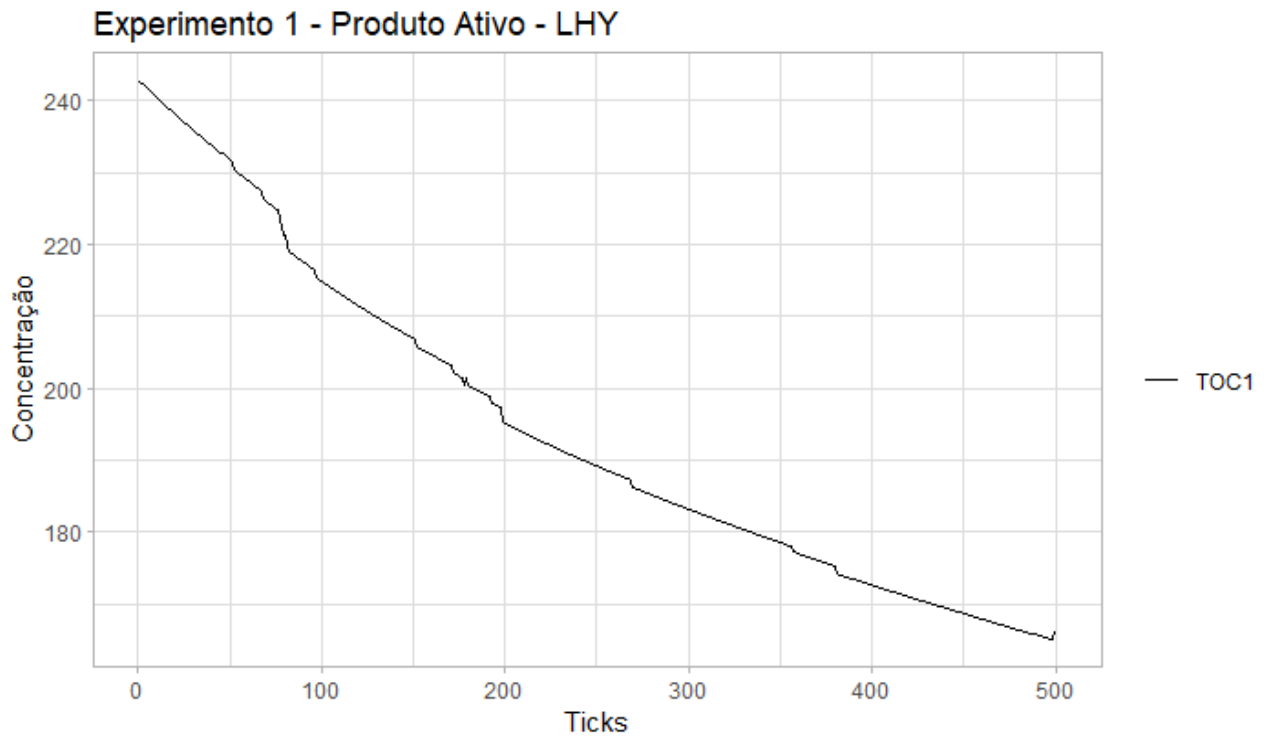


Figura 5.16: Concentração do gene LHY produzida

A Figura 5.17 mostra o gráfico do gene ZTL que é somente ativado pelo gene GI, em qualquer dos períodos definidos, sob incidência de luz. Portanto, o gráfico é somente ascendente.

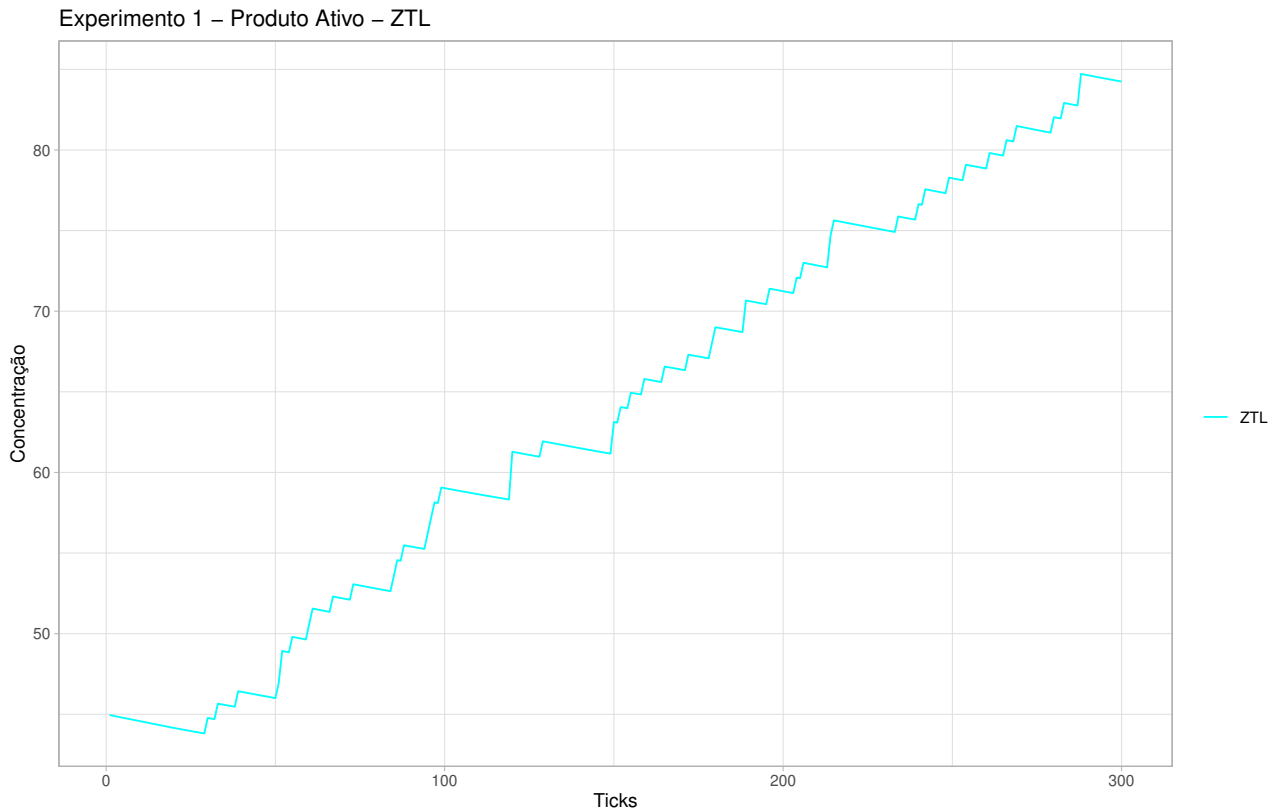
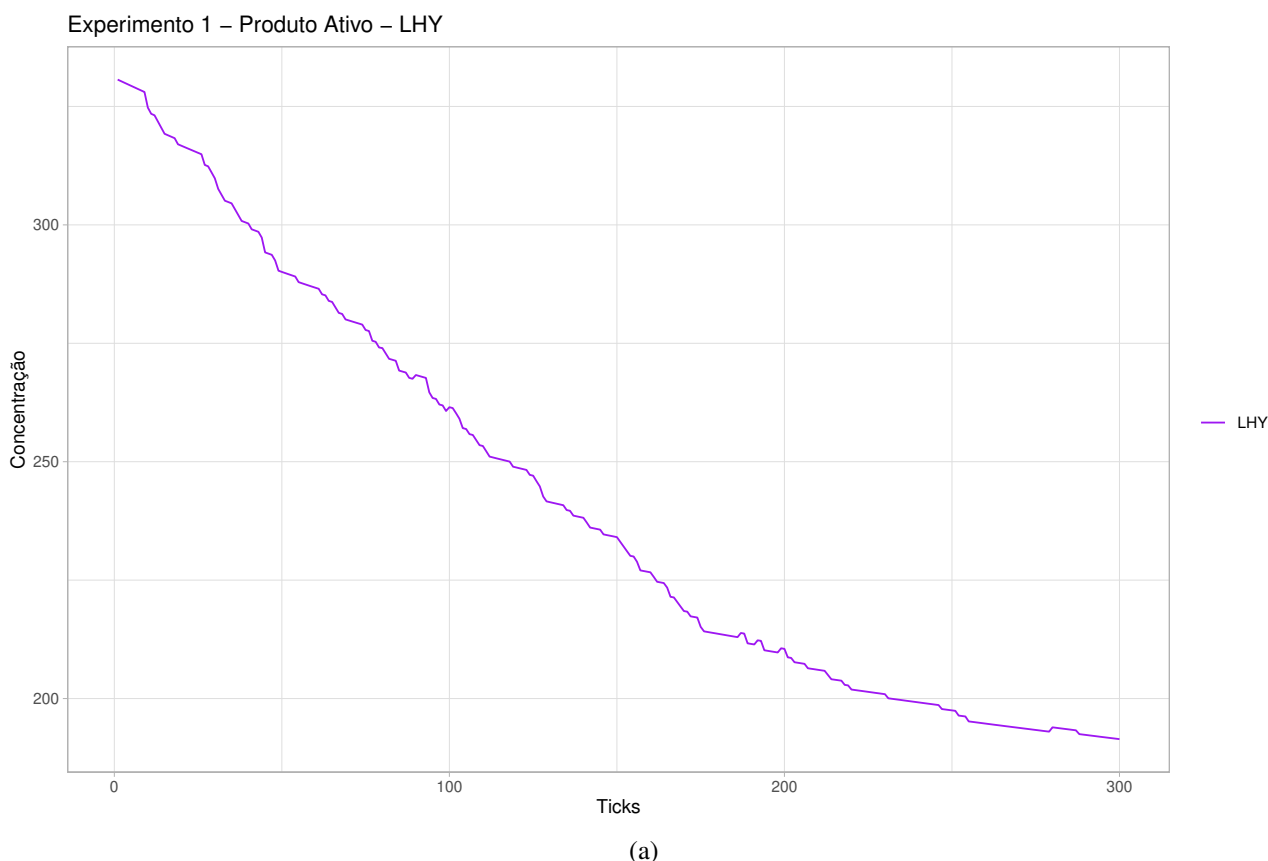


Figura 5.17: Concentração do gene ZTL produzida

A Figura 5.18a mostra o gráfico do gene LHY, que é inibido pelos genes PRR7, PRR9 e NI, em qualquer dos períodos definidos, sob influência da luz e somente é ativado pelo gene TOC1mod no período noturno (“nocturnal” - ticks 75 a 100), sem incidência de luz. Portanto, verifica-se que há somente alguns pequenos “picos” no gráfico descendente.



- É importante ressaltar que ocorre decremento do “produto ativo” durante toda a simulação e, por isso, há decaimento nos gráficos mesmo quando não ocorre inibição. Também considera-se importante ressaltar que todos os gráficos foram gerados com os valores do “produto ativo” de cada gene individualmente, sem calcular valores médios, para mostrar a capacidade de micro (individualidade) e flexibilidade do modelo. Apesar de alguns gráficos não mostrarem comportamento de alternância, pode-se perceber que mostram o comportamento adequado às configurações, ações, portas, incidência de luz e localização espacial.

5.3.3 Comparação do Modelo FMAS-GENERE com o modelo BioASF-BIOPAX

Como já mencionado, o Modelo da Rede Regulatória da planta *Arabidopsis* foi utilizado como estudo de caso para o framework desenvolvido nesta tese, o FMAS-GENERE. Para fins de comparação e estudos, foi realizado um paralelo entre os resultados do framework FMAS-GENERE e o resultados do modelo BioASF-BIOPAX.

O modelo de execução da simulação do BioASF-BIOPAX é representado como um rede na qual os nós são agentes que controlam a execução de seus interações biológicas correspondentes na rede biológica modelo. O modelo de execução de simulação impõe três tipos de restrições durante a execução de um modelo: (i) restrições definidas no BioPAX; (ii) restrições definidas por regras de simulação; e (iii) restrições definido por regras de análise. Portanto, o BioASF requer explicitamente entradas correspondentes [Haydarlou et al., 2016].

O framework é baseado nos princípios dos sistemas de eventos discretos (DES) e a teoria multiagente sistemas (MAS). Semelhante a um sistema discreto de eventos, os vários componentes da estrutura cooperam uns com os outros para executar uma função utilizando eventos. Cada componente gerencia um número de estados discretos e a transição de estado depende da ocorrência de eventos discretos [Haydarlou et al., 2016].

O comportamento apresentado no gráfico da Figura 5.18 mostra a oscilação na concentração dos genes PRR9, PRR7, NI, LHY e TOC1 durante um período equivalente a 5 dias, como resultado da simulação do framework BioASF-BIOPAX.

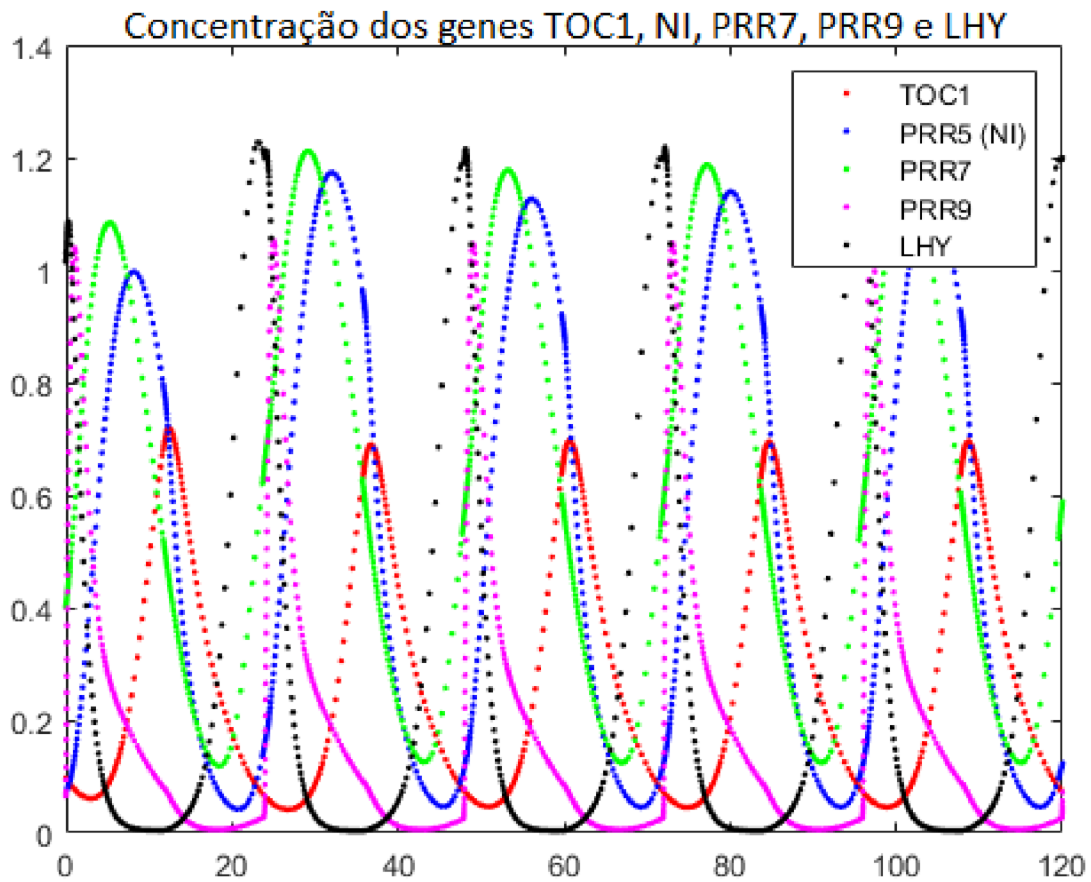


Figura 5.18: Concentrações dos genes PRR9, PRR7, NI, LHY e TOC1 resultantes da simulação realizada pela ferramenta BioPAX adaptado de [Pokhilko et al., 2010]

Já na Figura 5.19 observa-se o comportamento dos genes PPR9, PRR7 e NI na simulação do framework FMAS-GENERE durante um período de 500 ticks, que é equivalente aos 5 dias de simulação representados na Figura 5.18.

Esta Figura 5.19, como já explicado anteriormente nos resultados do framework FMAS-GENERE, mostra como PPR9 é ativado somente sob as condições de incidência de luz, energia de pelo menos 1 do gene LHY/CCA1 no sítio de ligação referente e encontrar-se no período matutino (“morning” - até tick 25), mas também sofre ação de inibição pelo gene TOC1 no período noturno (“nocturnal” -

dos ticks 75 a 100 ticks), o que gera os “picotados” no gráfico próximo ao final do período noturno. Repetindo-se o comportamento a cada dia.

O mesmo acontece para PRR7 na Figura 5.19, mostrando que o gene PRR7 é ativado somente sob as condições de incidência de luz, energia de pelo menos 1 do gene LHY/CCA1 no sítio de ligação referente e encontrar-se no período matutino (“morning” - até tick 25). Repetindo-se o comportamento a cada dia.

O comportamento de NI mostra que o gene é ativado somente sob as condições de incidência de luz, energia de pelo menos 1 do gene LHY/CCA1 no sítio de ligação referente e encontrar-se no período matutino (“morning” - até tick 25), apresenta um comportamento semelhante aos descritos acima e possui as mesmas configurações. Contudo, existe diferenças devido à localização espacial e à movimentação aleatória dos genes. Sendo que o comportamento repete-se a cada dia também.

Assim, pode-se perceber que apesar de não apresentar resultados semelhantes, o framework FMAS-GENERE apresentou comportamentos resultantes semelhantes aos do framework BioASF-BIOPAX, podendo-se observar a oscilação nas concentrações.

Porém, devido a condições diferentes que são consideradas no framework FMAS-GENERE, como espaço do grid em que os agentes movimentam-se, movimentos aleatórios dos agentes dentro do grid, tamanho dos agentes e das áreas que representam os sítios de ligação, configuração da quantidade de agentes que representam os genes, configuração da quantidade de agentes que interagem em cada sítio, também configuração dos motifs baseado nas configurações iniciais da rede da planta *Arabidopsis* e decaimento da energias dos genes no decorrer do tempo independentemente de qualquer outra condição, pode-se perceber alguns comportamentos diferentes do framework BioASF-BIOPAX, esta condição será explicada juntamente com a explicação da Figura 5.20

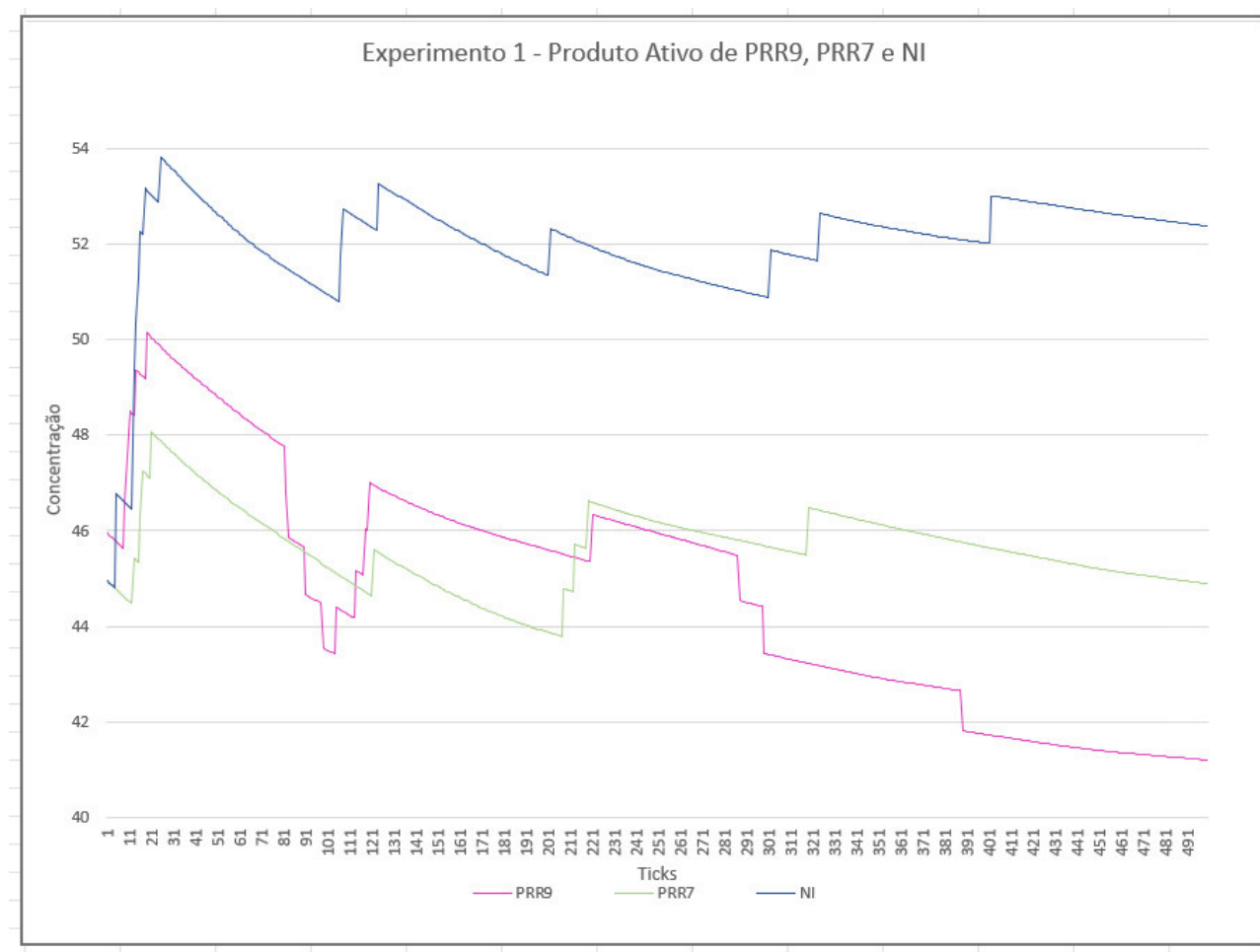


Figura 5.19: Concentração dos genes PPR9, PRR7 e NI

Ao comparar as Figuras 5.20 do framework FMAS-GENERE e 5.18, percebe-se que os resultados de LHY e TOC diferem do comportamento de LHY e TOC na simulação BioASF-BIOPAX.

A Figura 5.20 mostra o comportamento do gene TOC1, que é inibido pelo LHY/CCA1 no período vespertino (“vespertine” - ticks 50 a 75) sob incidência de luz, inibido pelo ZTL no período noturno (“nocturnal” - ticks 75 a 100) e ativado pelo GI no período noturno (“nocturnal” - ticks 75 a 100), sob remanescente de luz. E assim, devido ao fato de ser inibido em mais de um sítio de ligação, devido mais de uma condição e existir o decaimento da energia de todos genes independentemente de qualquer condição pré-estabelecida, na simulação do do framework FMAS-GENERE observa-se um comportamento de decaimento.

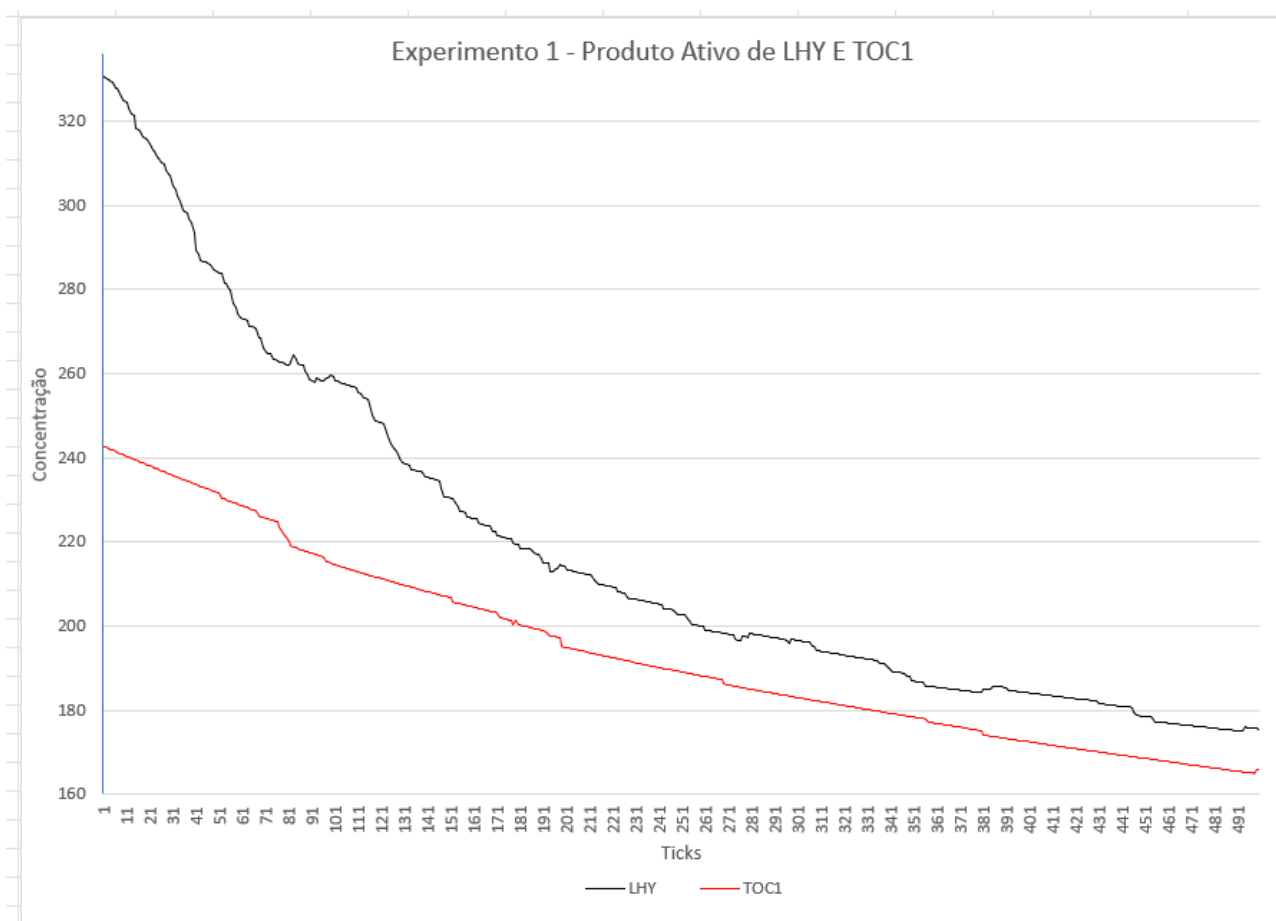


Figura 5.20: Concentração dos genes LHY e TOC1

A Figura 5.20 também mostra o gráfico do gene LHY, que é inibido pelos genes PRR7, PRR9 e NI, em qualquer dos períodos definidos, sob influência da luz e somente é ativado pelo gene TOC1mod no período noturno (“nocturnal” - ticks 75 a 100), sem incidência de luz. Portanto, verifica-se que há somente alguns pequenos “picos” no gráfico descendente, diferentemente do que é apresentado no comportamento da simulação do BioASF-BIOPAX, porém explicado devidos as diferenças já citadas entre os frameworks, mostrando-se assim o nível de detalhamento informações diferentes que cada framework pode trazer com suas particularidades.

6 CONCLUSÕES E TRABALHOS FUTUROS

Este trabalho teve com objetivo geral o desenvolvimento de um simulador multiagente para GRN, visto que na literatura ainda há uma demanda por ferramenta que permita a modelagem e simulação de GRN de maneira mais ampla e flexível, sendo possível visualizar aspectos a nível micro (de cada gene) e macro (a nível do ambiente como um todo). Para tanto, inicialmente é apresentado o embasamento teórico da área e definido um modelo baseado na estrutura de “motifs” para embasar o framework proposto. De forma a validar a ferramenta desenvolvida, três estudos de caso são apresentados, com níveis de complexidade (modelagem, restrições e variáveis).

Após realizar o desenvolvimento do framework e adaptar as configurações e condições de acordo com a evolução do modelo devido às validações com os estudos de caso, percebe-se que em todos os estudos de caso o framework apresentaram comportamentos semelhantes aos resultados apresentados em trabalhos prévios. Contudo, a modelagem, implementação e simulação no framework proposto apresentou flexibilidade ao mostrar que pode ser facilmente configurado de acordo com a GRN em estudo, pois utiliza arquivos de entrada configurados de forma simples, baseados em estruturas de “motifs”. O framework também apresentou sua capacidade de representar cada gene/agente individualmente, o que não é possível em modelos matemáticos, o que traz riqueza em detalhes de informações.

O framework FMAS-GENERE apresenta como principal vantagem a possibilidade de realizar a simulação de qualquer rede regulatória que possa ser configurada de acordo com o padrão de estrutura de motifs e de acordo com o padrão estabelecido para os arquivos de entrada do framework, mas também apresenta como vantagem o fato de não precisar configurar variáveis para simulação baseada em ODE, já que é um framework baseado na interação entre agentes, além de não precisar de nenhuma simulação prévia de qualquer tipo para gerar ou tratar os dados de entrada.

O framework FMAS-GENERE apresenta como principais diferenças em relação ao framework e outras ferramentas características como espaço do grid em que os agentes movimentam-se, movimentos aleatórios dos agentes dentro do grid, tamanho dos agentes e das áreas que representam os sítios de ligação, configuração da quantidade de agentes que representam os genes, configuração da quantidade de agentes que interagem em cada sítio, também configuração dos motifs baseado nas configurações iniciais da rede a ser simulada e decaimento de energia dos genes no decorrer do tempo independentemente de qualquer outra condição.

Observa-se o fato de não ter um método ou características compatíveis que possam ser utilizados como embasamento para calibração inicial dos dados de entrada como uma desvantagem do framework FMAS-GENERE frente a outras ferramentas que utilizam ODEs ou resultados de experimentos *in vitro*.

O framework também pode ser facilmente adaptado para retornar como resultados as variáveis e informações desejadas de acordo com a GRN e o estudo de caso. Sendo assim, pode-se concluir que o modelo atendeu seu objetivo geral que é o desenvolvimento de um simulador multiagente para GRN contendo abstração dos processos bioquímicos subjacentes e que simula a interação entre o TF

(*Transcription Factor*) e um gene alvo em GRNs genéricas.

Também são resultados desta tese de doutorado as seguintes publicações:

- Revisão Sistemática para o Desenvolvimento de um Ambiente de Simulação Multiagente para Vias Regulatórias [Agostinho et al., 2018a] - REVISTA DE TECNOLOGIAS (OURINHOS), v. 11, p. 39-50, 2018;
- Um modelo para simulação de formação de membranas e micelas [Agostinho et al., 2018b] - Workshop-Escola de Sistema de Agentes, seus Ambientes e aplicações, Fortaleza, 2018;
- Development of a Multiagent Simulator to Genetic Regulatory [Agostinho et al., 2019] - Latin-American Workshop on Computational Neuroscience, 2019, São João del Rey / MG. LAWCN 2019 ;
- Development of a Multiagent Simulator to Genetic Regulatory Networks [Agostinho et al., 2020b] - 18th International Conference on Practical Applications of Agents and Multi-Agent Systems, 2020, L'Aquila / Italia. PAAMS 2020;
- A Systematic Review to Multiagent Systems and Regulatory Networks [Agostinho et al., 2020a] - 17th International Conference on Distributed Computing and Artificial Intelligence, 2020, L'Aquila / Italia. DCAI 2020;
- Uma revisão sistemática para modelagem de redes regulatórias genéticas utilizando sistemas multiagente [Agostinho et al., 2022] - REVISTA ELETRÔNICA ARGENTINA-BRASIL DE TECNOLOGIAS DA INFORMAÇÃO E DA COMUNICAÇÃO, v. 15, p. 1-19, 2022.

Assim, pode-se concluir, do ponto de vista científico, que a utilização de MAS para modelagem e simulação de GRN é possível, e que a visualização de aspectos desejados a nível micro/macro é possível se der realizada, permitindo novas análises e formas de entendimento para as pesquisas na área. Do ponto de vista tecnológico, o desenvolvimento de uma ferramenta que permite a fácil utilização (a partir de arquivos de entrada configurados) também é muito interessante, pois pesquisadores das áreas das ciências biológicas também podem fazer uso de forma mais ampla.

Como trabalhos futuros visumbra-se:

- Desenvolver um software que possua interface e possa ser utilizado para configurar as taxas de deicamento e que calcule as quantidades de genes de acordo com a rede a ser simulada e o grid configurado para então gerar os arquivos de entrada para o framework.
- Realizar a geração automática dos arquivos de saída independentemente da rede a ser configurada e modelada.
- Aprimorar o modelo para que possua mais opções de configuração e possua uma integração com “R” ou outra linguagem/aplicativo para geração automática dos gráficos.

Referências

- N. B. Agostinho, K. S. Machado, and A. V. Werhli. Inference of regulatory networks with mcmc sampler guided by mutual information. *Proceedings of ACM SAC*, 2017.
- N.B. Agostinho, A. V. Werhli, and D. F. Adamatti. Revisão sistemática para o desenvolvimento de um ambiente de simulação multiagente para vias regulatórias. *RETEC, Ourinhos*, 11:39–50, 2018a.
- N.B. Agostinho, A. V. Werhli, and D. F. Adamatti. Development of a multiagent simulator to genetic regulatory. *II Latin American Workshop on Computational Neuroscience*, 2019.
- N.B. Agostinho, A. V. Werhli, and D. F. Adamatti. A systematic review to multiagent systems and regulatory networks. *Part of the Advances in Intelligent Systems and Computing book series*, 2020a.
- N.B. Agostinho, A. V. Werhli, and D. F. Adamatti. Development of a multiagent simulator to genetic regulatory networks. *Part of the Advances in Intelligent Systems and Computing book series*, 2020b.
- N.B. Agostinho, A. V. Werhli, and D. F. Adamatti. Uma revisão sistemática para modelagem de redes regulatórias genética utilizando sistemas multiagente. *Revista Eletrônica Argentina-Brasil de Tecnologias da Informação e da Comunicação*, 2022.
- Nilzair B. Agostinho, Karina S. Machado, and Adriano V. Werhli. Inference of regulatory networks with a convergence improved mcmc sampler. *BMC Bioinformatics*, 2015.
- Nilzair B. Agostinho, Karina S. Machado, and Adriano V. Werhli. Um modelo para simulação de formação de membranas e micelas. In *Workshop-Escola de Sistema de Agentes, seus Ambientes e aplicações*, Fortaleza, 2018b.
- Uri Alon. Network motifs: theory and experimental approaches. *Nature Reviews Genetics*, 2007.
- Roberto Barbuti, Roberta Gori, Paolo Milazzo, and Lucia Nasti. A survey of gene regulatory networks modelling methods: from differential equations, to boolean and qualitative bioinspired models. *Journal of Membrane Computing*, 2(3):207–226, 2020. ISSN 2523-8914. doi: 10.1007/s41965-020-00046-y. URL <https://doi.org/10.1007/s41965-020-00046-y>.
- B. N. Brown, I. M. Price, and F. R. Toapanta. An agentbased model of inflammation and fibrosis following particulate exposure in the lung. *Mathematical Biosciences*, 2011.
- F. Chiacchio, M. Pennisi, G. Russo, S. Motta, and F. Pappalardo. Agent-based modeling of the immune system: Netlogo, a promising framework. *Hindawi Publishing Corporation BioMed Research International*, 2014.
- D. Dreau, D. Stanimirov, T. Carmichael, and M. Hadzikadic. An agent-based model of solid tumor progression. *International Conference on Bioinformatics and Computational Biology*, 2009.

- Ezer. Homotypic clusters of transcription factor binding sites: A model system for understanding the physical mechanics of gene expression. *Comput Struct Biotechnol J.*, 2014.
- N. Friedman, M. Linial, I. Nachman, and D. & Peter. Using bayesian networks to analyze expression data. *Journal of Computational Biology*, 2000.
- T.F. Galvao and M.G. Pereira. Etapas de busca e seleção de artigos em revisões sistemáticas da literatura. *Epidemiologia e Servico de Saude*, 23:369–371, 2014.
- A. Ghazikhani, T.M.R. Akbarzadeh, and R. Monsefi. Genetic regulatory network inference using recurrent neural networks trained by a multi agent system. *International eConference on Computer and Knowledge Engineering (ICCKE)*, 2011.
- C Gibas and P Jambeck. Developing bioinformatics computer skills. *Yale Journal of Biology and Medicine.*, 2002.
- Chad M Glen, Melissa L Kemp, and Eberhard O Voit. Agent-based modeling of morphogenetic systems: Advantages and challenges. *PLoS Computational Biology*, 15(3), mar 2019. doi: 10.1371/journal.pcbi.1006577.
- M.L. Guerriero, A. Pokhilko, A.P Fernandez, K.J Halliday, A.J. Millar, and J. Hillston. Stochastic properties of the plant circadian clock journal of the royal society interface. *Journal of the Royal Society Interface*, 2012.
- R. Haydarlou, A. Jacobsen, N. Bonzanni, K. A. Feenstra, S. Abeln, and J. Heringa. Bioasf: a framework for automatically generating executable pathway models specified in biopax. *Bioinformatics.*, 32:i60–i69, 2016.
- T.J Hearn, M.C.M Ruiz, S.M. Abdul-Awal, R. Wimalasekera, C.R. Stanton, M.J. Haydon, F.L. Theodoulou, M.A. Hannah, and A.A.R Webb. Big regulates dynamic adjustment of circadian period in arabidopsis thaliana. *Plant Physiol.*, 2018.
- N. R. Jennings. Coordination techniques for dai. in ohare, greg jennings, nicholas (eds.). *Foundations of distributed artificial intelligence.*, 1996.
- K.A. Johnson and R.S. Goody. The original michaelis constant: Translation of the 1913 michaelis-menten paper. *Biochemistry*, 2011.
- S. Khan, R. Makkena, F. McGeary, K. Decker, W. Gillis, and C. Schmidt. A multi-agent system for the quantitative simulation of biological networks. *AAMAS*, 2003.
- J. Liu, Y. Z. C Chi, and C. Zhu. A dynamic multiagent genetic algorithm for gene regulatory network reconstruction based on fuzzy cognitive maps. *IEEE TRANSACTIONS ON FUZZY SYSTEMS*, 24, 2016.

- F. M Lopes. Redes complexas de expressao genica: sintese, identificacao, analise e aplicacoes. Master's thesis, Universidade de Sao Paulo, 2011.
- D.C.B Mariano, C Leite, L.H.S Santos, R.E.O Rocha, and R.C Melo-Minardi. A guide to performing systematic literature reviews in bioinformatics. Technical Report RT.DCC.002/2017, Universidade Federal de Minas Gerais, 2017.
- M. Pennisi, A. M. Rajput, L. Toldo, and F. Pappalardo. Agent based modeling of treg-teff cross regulation in relapsingremitting multiple sclerosis. *BMC Bioinformatics*, 2013.
- D.H Pham. Multi-agent based simulation of large random boolean network. *MSC-LES*, 2008.
- A Pokhilko, S.K. Hodge, K. Stratford, K. Knox, K.D. Edwards, A.W. Thomson, and T. Mizuno. Data assimilation constrains new connections and components in a complex, eukaryotic circadian clock model. *Molecular Systems Biology*, 2010.
- A Pokhilko, A.P. Fernandez, K.D. Edwards, M.M. Southern, K.J. Halliday, and A.J. Millar. The clock gene circuit in arabidopsis includes a repressilator with additional feedback loops. *molecular systems biology*. *Molecular System Biology*, 8, 2012.
- A.P. Sanfilippo, J.N. Haack, J.E McDermott, S.L. Stevens, and M.P Stenzel-Poore. Modeling emergence in neuroprotective regulatory networks. *International Conference on Complex Sciences - springer*, pages 291–302, 2012.
- S.L. Schreiber. Small molecules: the missing link in the central dogma. *Nature Chemical Biology*, 2005.
- Z. Z. Shawn Martin, A. Martino, and J.L. Faulon. Boolean dynamics of genetic regulatory networks inferred from microarray time series data. *Bioinformatics*, 2005.
- I. Shmulevich, E.R. Dougherty, S. Kim, and W. Zhang. Probabilistic boolean networks: a rule-based uncertainty model for gene regulatory networks. *Bioinformatics*, 2002.
- S. Tisue and U. Wilensky. Netlogo: A simple environment for modeling complexity. *Presented at the International Conference on Complex Systems, Boston*, 2004.
- A. V. Werhli. *Reconstruction of gene regulatory networks from postgenomic data*. PhD thesis, School of Informatics University of Edinburgh, 2007.
- A. V. Werhli and D. Husmeier. Reconstructing gene regulatory networks with bayesian networks by combining expression data with multiple sources of prior knowledge. *Statistical Applications in Genetics and Molecular Biology*, 2007.
- P. Werlang, M.Q. Fagundes, D.F. Adamatti, K.S. Machado, A. von Groll, P.E.A Silva, and A.V Werhli. Multi-agent-based simulation of mycobacterium tuberculosis growth. in: Alam s., parunak h. (eds) multi-agent-based simulation. *International Workshop on Multi-Agent Systems and Agent-Based Simulation MABS 2013: Multi-Agent-Based Simulation XIV*, 8235:131–142, 2014.

- M. Wooldridge and N. R. Jennings. Intelligent agents: Theory and practice. *The Knowledge Engineering Review*, 1995.
- Hulin Wu, Tao Lu, Hongqi Xue, and Hua Liang. Sparse additive ordinary differential equations for dynamic gene regulatory network modeling. *Journal of the American Statistical Association*, 109(506):700–716, 2014. doi: 10.1080/01621459.2013.859617. URL <https://doi.org/10.1080/01621459.2013.859617>. PMID: 25061254.
- T. Yang and Y.F. Sun. The reconstruction of gene regulatory network based on multi-agent system by fusing multiple data sources. *2011 IEEE International Conference on Computer Science and Automation Engineering*, 4, 2011.