Cabaret: A Suite of Characterized Discrete Models for Benchmarking Systems Biology Tools

by

Handa Ding

B.S., Jiangxi Science and Technology Normal University, 2013

M.S., University of Texas at Dallas, 2016

Submitted to the Graduate Faculty of the

Swanson School of Engineering in partial fulfillment

of the requirements for the degree of

Master of Science in Electrical and Computer Engineering

University of Pittsburgh

2019

UNIVERSITY OF PITTSBURGH

SWANSON SCHOOL OF ENGINEERING

This thesis was presented

by

Handa Ding

It was defended on

November 14, 2019

and approved by

Natasa Miskov-Zivanov, Ph.D., Assistant Professor
Department of Electrical and Computer Engineering

Jun Yang, Ph.D., Professor
Department of Electrical and Computer Engineering

Samuel Dickerson, Ph.D., Assistant Professor
Department of Electrical and Computer Engineering

Cheryl Telmer, Ph.D., Senior Molecular Biologist
Carnegie Mellon University

Thesis Advisor: Natasa Miskov-Zivanov, Ph.D., Assistant Professor
Department of Electrical and Computer Engineering

Cabaret: A Suite of Characterized Discrete Models for Benchmarking Systems Biology Tools

Handa Ding, M.S.

University of Pittsburgh, 2019

Computational modeling tools have a great potential to increase our understanding of complex biological systems. With the significant increase in publicly available biomedical data, systems biology methods have advanced, resulting in numerous frameworks that are accessible for the whole community. However, with many different methods proposed to solve computational biology problems, there is a tremendous need for standards to measure capabilities of the tools. Accordingly, assessing and deciding which tools to use to address specific questions and problems is often a considerable challenge. One solution to address this challenge is to follow the practice of other fields, such as computer engineering, and create benchmarks. In biology, a suite of models that are evaluated using typical measures would allow researchers to compare the performance among different tools in an unbiased fashion. Thus, we propose CABARET, a **C**haracterized **A**ssembly of **B**enchmarks for **A**utomation, **R**eproducibility and **E**valuation of **T**ools in biology.

Our benchmark suite (CABARET) will provide a set of models and analysis methods for biological modelers to comprehensively evaluate their tools. In this thesis, we selected seven discrete cell signaling and gene regulation network models of immune system and cancer cells. These models are then simulated both deterministically and stochastically to explore their steady states and transient responses. Using the same simulation approach and analysis methods on all models allows for standardized measures and reporting of the features of the models and results in

a well characterized set of benchmark models. For this characterization, feedback loop, fan-in and fan-out cone analyses, as well as the analysis of paths between inputs and outputs, and additional features of scenarios specific for modeled systems have been documented. We believe these models can serve as standardized, calibrated, benchmarks to evaluate future tools developed by the biology community.

Keywords: Biological Systems, Boolean Model, Cancer Systems, Benchmarking Characteristics, Logic, CABARET, Python, Automation, Tools.

Table of Contents

List of Tables

List of Figures

Preface and Acknowledgement

First and foremost, I would like to thank my supervisor Professor Natasa Miskov-Zivanov for her generous support. You have always been very helpful and patient when unexpected challenges happen. During my study and research, I have benefited tremendously from your knowledgeable expertise and guidance in both electrical engineering and bioengineering. Furthermore, colleagues from MeLoDy Lab, Dr. Cheryl Telmer, Dr. Kara Bocan, Emilee Holtzapple, Khaled Sayed, Adam Butchy, Gavin Zhou, Yasmine Ahmed, Niteesh Sundaram, Stefan Andjekovic, and Casey Hansen for their selfless guidance and help! Also, I would like to thank the ECE department of University of Pittsburgh for their financial support and teaching assistant opportunities through my graduate career. Additionally, special thanks to Defense Advanced Research Projects Agency (DARPA) for providing such a remarkable research platform and their partial funding as well.

Of course, and most importantly, I would like to thank my whole family for their unremitting support and many years of encouragement. I am deeply touched and appreciative, and this work is dedicated to all of you. Finally, like any thousand mile journey, I believe my work CABARET will be the first step, and eventually it will benefit for all mankind in the future.

1.0 Introduction

Benchmarking is defined as evaluating or checking by comparison with a standard, so the differences can be used to identify gaps in a method [1] in order to achieve a strategic advantage. The terminology first originated in the entrepreneurial community as a common strategy: when you have no idea what to do, check what your competitors are doing instead of creating a new product [2]. In other words, benchmarking is a procedure for acquiring a standard for performance. Benchmarks are the "what", and benchmarking is the "how". Creating and updating solid benchmarks significantly lower operating costs and improve efficiency within limited resources [3].

Currently, benchmarking methodology is not only well adopted in many business areas, such as corporate management, where the benchmark approach helps to understand fundamental relationships in financial market [4]. It is also used in other fields such as higher education, where benchmarks lead to institutionalized assessment practice – a unique methodology can improve learning outcomes of undergraduate business program [5]. Benchmarking also plays a critical role in computer architecture, where standardized assessments of, for example, CPU, hard drive, or power consumption, are used to help researchers compare computer system performance across different platforms. Computational biology methods also require rigorous benchmarks in order to compare performance.

In experimental biology there are many standards and calibrating compounds, and there are nomenclature and data standard in genomics, however there is a lack of standard for computational modeling. There has been a large increase in the number and types of models being created, but it is difficult to compare approaches to select the method most appropriate for the

biological question. In the past few decades, the number of systems biology publications has grown rapidly. Correspondingly, the amount of experimental data has increased, and by computationally modeling biological systems, we can better understand the relationship from the cellular level up to the macroscopic level [6]. Usually, there are a number of solutions for modeling biological systems including the Monte-Carlo method [7], ordinary differential equations (ODE) [8], rule-based modeling approach [9], and Boolean network or logical modeling approach. These methods can provide both efficient and accurate analysis. As we know, the biology community has already developed many methods and software packages to address the dynamic modeling issue. However, there is lack of suitable benchmark suites which allow a fair and systematic evaluation of these contributions. Owing to the extremely complex nature of biological systems, creating a gold standard for the crucial list of model characteristics that can suitably highlight features of model building and analysis tools becomes a necessary and urgent task.

In this work, we focus on discrete logical models of biological signaling networks. A Boolean model is a special case of discrete model that has been used to mimic the ON (logic 1) and OFF (logic 0) state of each element. Accordingly, there are a few analytical methods in biological modeling. For example, sensitivity analysis plays a critical role in measuring system robustness and adaptability under perturbations [10]. Applying both static and dynamic sensitivity analysis can provide various different perspectives based on different applications—the former one focuses on uniform state distribution, the latter one uses distribution from stochastic simulation results [10].

In this work, we present seven benchmarks, seven discrete logical models: Albert's ABA plant model [11] and large granular leukemia (LGL) model [12], Sahin's breast cancer model [13], Bauer's angiogenesis model [14], naïve Tcell differentiation model [15], macrophage model [16],

and pancreatic cancer cell (PCC) model [17]. The bio-benchmark suite uses standardized spreadsheet representation format [18] for all models, from which both graphical representation of benchmarks and executable models can be generated. The standardized and systematic description of the characteristics of the static model networks and dynamic model behavior will enable reliable evaluation of systems biology tools, and therefore, lead to higher confidence in findings provided by these tools.

2.0 Background

2.1 Existing benchmark suites in various fields

Successful benchmarks provide a standard performance evaluation between the new product and the existing reference. In other words, technological gaps can be identified, thus further improvement can be done. Hence, benchmarking methodology has been widely used in various fields such as business and technology as mentioned in the Chapter 1.0. In this paper, we learn more lessons and acquire experience from the benchmark suites in electrical and computer engineering and I believe their processes will lead us to establish better benchmarks for the biology community

2.1.1 SPEC benchmark suites

The well-known industry SPEC (System Performance Evaluation Corporation) tool was invented in 1988 by a small group of workstation vendors who realized that a set of standardized performance tests for computer systems was desperately needed for the current market. Nowadays, SPEC has successfully become one of the best performance evaluation entities and it contains over 60 members including hardware and software vendors, universities, and researchers. Specifically, the SPEC benchmark suite is an umbrella structure that has diverse subcommittees, which contains evaluation kits for CPU (central processing units), Cloud, Graphics, Workstation, High performance computing, Power, Storage.

Among all kits, SPEC CPU® is one of the oldest groups within SPEC OSG (Open Systems Group), of which the most updated version *CPU 2017* owns 43 benchmarks, categorized into 4 groups: SPECrate®2017-Integer, SPECspeed®2017-Integer, SPECrate®2017-Floating Point, and SPECspeed® 2017-Floating Point[19]. For SPECrate®, the calculated metrics are based on equation 2-1:

$$\frac{time\ on\ a\ reference\ machine}{time\ on\ the\ SUT} \qquad 2\text{-}1$$

Naturally, the higher scores mean that less time is needed for SUT (system under test). As for SPECspeed®, the tester chooses the number of parallel copies to run, thus the performance is given by equation 2-2:

$$\frac{number\ of\ copies * (time\ on\ a\ reference\ machine)}{time\ on\ the\ SUT} \qquad 2\text{-}2$$

And the higher scores indicate more work can be done per unit of time. Based on those metrics, the readers or consumers can easily compare the performance metrics between the new product and existing benchmarks, allowing them to make an evaluation on processor, memory and compilers in an unbiased fashion.

2.1.2 Embedded Microprocessor Benchmarks Consortium (EEBC).

EEBC was established in 1997 and focuses on creating the standard benchmarks for hardware and software in embedded systems [20]. The member companies can use benchmark suites to evaluate the performance of their own devices either for business purpose or further improvement in next-generation product. Currently, the benchmark contains several suites toward different applications, such as: Ultra-Low Power and Internet of Things, Heterogeneous Compute, Single-core processor performance, Multi-core processor performance, Phone and tablet [21]. For example, CoreMark of Single-core processor performance group can measure the performance of CPU used in embedded systems, and the evaluation score is based on the performance of list, matrix, and state machine processing [22].

2.1.3 Lessons from SPEC and EEBC

For both SPEC and EEBC, each benchmark suites measures the performance of hardware based on several algorithms. As an example, in SPECrate® CPU 2017 Floating Point Rate Result for ASUS RS700-E9 Server System, 13 different benchmarks are selected to measure the overall floating-point performance [23], and the final metrics gives an overall fair evaluation for the Server System. Similarly, in modeling biological systems by using new different tools, if the ground truth characteristics of the model has been provided, researchers can compare characteristics and make a much more unbiased decision among various tools. Repositories that contain automatically generated models, such as BioModels [24], would benefit from a standardized benchmark suite. Different modeling results of the same bio-medical system be compared to the benchmarks, so that researcher can select the best tool in an unbiased fashion.

2.2 Discrete and logical modeling approaches

To study the increasing complexity of cell microenvironment, typically a large biological signaling network, discrete element-based modeling has advantage over the traditional types (such as ODE, rule-based, etc.) [25]. Simulating a large complex system usually takes from hours to days with ODE solvers. Hence, creating a biological model at a higher level of abstraction by implementing finite set of discrete levels and neglecting a few precise mechanism details, allows for capturing direct or indirect system dynamics, while significantly saving runtime [26]. Boolean logic as the special case in discrete logic is the modeling approach of choice for all the benchmarks we discuss in this work. Proposed by Kauffman in 1969, Boolean network concept was implemented for description of a biological system to observe and measure biosynthesis [27], where each element has two possible states, either ON or OFF. Also, logic rules are associated with a subset of elements and are used to update elements' states at each time step according to the states of their regulators. There are two types of regulations, positive and negative, and they are represented graphically as regular and blunt arrows, respectively, as shown in Figure 1. Specifically, from Figure 1, element A positively regulates element B (A→B); and element B negatively regulates element C (B —| C); also, element C negatively regulates element A (C —| A).

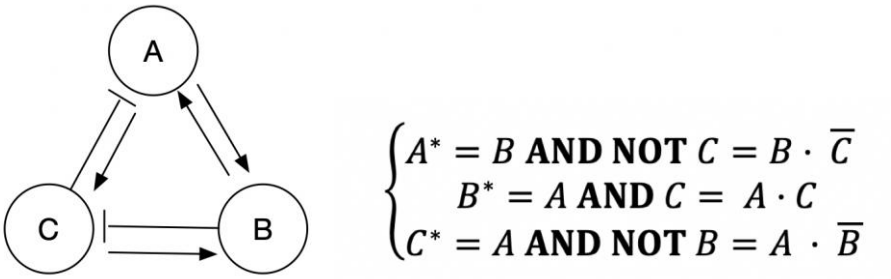

Figure 1: Toy example of a logical model

If we include all the regulation information for each element, where each element could be either positively or negatively regulated by other elements, then the next state of every element is computed as a function of the previous state of its regulators. More formally, for any node y_i in the model that consists of *m* elements, the regulatory function (or, update function) can be written as in equation 2-3:

$$y_i^* = f_i(y_1, y_2, \dots, y_m) \qquad 2\text{-}3$$

Where y_i^* represents the next step logic state of node y_i, and y_1, y_2,…, y_m represent the rest of elements state before updating y_i. We show an example of update functions written as Boolean expressions for the toy model in Figure 1. Using these update functions, we can compute state transitions of the system, starting from a given initial state, until the system reaches steady state, or for a given number of time steps.

3.0 Benchmark Characteristics

Similar to the benchmark suites we discussed in Chapter 2 [23], we created comprehensive metrics to measure the performance of tools for modeling and analysis in systems biology. Accordingly, a suite of benchmarks, **C**haracterized **A**ssembly of **B**enchmarks for **A**utomation, **R**eproducibility and **E**valuation of **T**ools in Biology (CABARET), and the standardized list of benchmark characteristics and metrics is proposed in this thesis. This list includes the following three groups of characteristics.

1) **Standardized representation format** for discrete models of biological networks: We use a standardized spreadsheet format as the simulation input, the spreadsheet describes the biological system in many perspectives, such as: Element Type, Element IDs, Cell Lines, Cell Type, Organism, Tissue Type, Location, Location ID, Variable, Positive Regulators, Negative Regulators, Model Input (I) / Output (O), Levels. For example, breast cancer model is illustrated in the standard spreadsheet format shown in Figure 2, which acts as a comprehensive recipe for bio-modeling [18].

Element Ty	Element IDs	Cell Line	Cell Ty	Organi	Tissue T	Locatio	Location ID	Variable	Positive Regulators	Negative Regulators	Input (I) /	Levi
protein	P00533	HCC1954, SK-BR-3,	Breast	Human	Breast	cytoplasm	GO:0005737	ERBB1pn_cytoBR	EGFpn_cytoBR		input/output	2
protein	P04626	HCC1954, SK-BR-3,	Breast	Human	Breast	cytoplasm	GO:0005737	ERBB2pn_cytoBR	EGFpn_cytoBR		input/output	2
protein	P21860	HCC1954, SK-BR-3,	Breast	Human	Breast	cytoplasm	GO:0005737	ERBB3pn_cytoBR	EGFpn_cytoBR		input/output	2
protein comple	P00533,P04626	HCC1954, SK-BR-3,	Breast	Human	Breast	plasma men	GO:0005886	ERBB12pf_memBR	(ERBB1pn_cytoBR,ERBB2pn_cytoBR)		input/output	2
protein comple	P00533,P21860	HCC1954, SK-BR-3,	Breast	Human	Breast	plasma men	GO:0005886	ERBB13pf_memBR	(ERBB1pn_cytoBR,ERBB3pn_cytoBR)		input/output	2
protein comple	P04626,P21860	HCC1954, SK-BR-3,	Breast	Human	Breast	plasma men	GO:0005886	ERBB23pf_memBR	(ERBB2pn_cytoBR,ERBB3pn_cytoBR)		input/output	2
protein	Q6QBS2	HCC1954, SK-BR-3,	Breast	Human	Breast	cytoplasm	GO:0005737	EGFpn_cytoBR			input	2
protein	P08069	HCC1954, SK-BR-3,	Breast	Human	Breast	plasma men	GO:0005886	IGF1Rpn_memBR	ERALPHApn_cytoBR,AKT1pn_cytoBR	ERBB23pf_memBR	input/output	2
protein	GC06P151656	HCC1954, SK-BR-3,	Breast	Human	Breast	cytoplasm	GO:0005737	ERALPHApn_cytoBR	AKT1pn_cytoBR,MEK1pn_cytoBR		input/output	2
protein	GC08P127735	HCC1954, SK-BR-3,	Breast	Human	Breast	cytoplasm	GO:0005737	CMYCpn_cytoBR	pn_cytoBR,MEK1pn_cytoBR,ERALPHApn_cytoBR		input/output	2
protein	P31749	HCC1954, SK-BR-3,	Breast	Human	Breast	cytoplasm	GO:0005737	AKT1pn_cytoBR	pf_memBR,ERBB13pf_memBR,ERBB23pf_memBR,IGF1Rpn_memBR		input/output	2
protein family	Q02750,P36507	HCC1954, SK-BR-3,	Breast	Human	Breast	cytoplasm	GO:0005737	MEK1pn_cytoBR	pf_memBR,ERBB13pf_memBR,ERBB23pf_memBR,IGF1Rpn_memBR		input/output	2
protein	P24941	HCC1954, SK-BR-3,	Breast	Human	Breast	cytoplasm	GO:0005737	CDK2pn_cytoBR	CYCLINE1pn_cytoBR	(P21pn_cytoBR,P27pn_cytoBR)	input/output	2
protein	P11802	HCC1954, SK-BR-3,	Breast	Human	Breast	cytoplasm	GO:0005737	CDK4pn_cytoBR	CYCLIND1pn_cytoBR	(P21pn_cytoBR,P27pn_cytoBR)	input/output	2
protein	Q00534	HCC1954, SK-BR-3,	Breast	Human	Breast	cytoplasm	GO:0005737	CDK6pn_cytoBR	CYCLIND1pn_cytoBR		input/output	2
protein	P24385	HCC1954, SK-BR-3,	Breast	Human	Breast	cytoplasm	GO:0005737	CYCLIND1pn_cytoBR	(ERALPHApn_cytoBR,CMYCpn_cytoBR,AKT1pn_cytoBR),(ERALPHApn_(		input/output	2
protein	P24864	HCC1954, SK-BR-3,	Breast	Human	Breast	cytoplasm	GO:0005737	CYCLINE1pn_cytoBR	CMYCpn_cytoBR		input/output	2
protein	P38936	HCC1954, SK-BR-3,	Breast	Human	Breast	cytoplasm	GO:0005737	P21pn_cytoBR	ERALPHApn_cytoBR	(AKT1pn_cytoBR,CMYCpn_cyto	input/output	2
protein	GC12P012716	HCC1954, SK-BR-3,	Breast	Human	Breast	cytoplasm	GO:0005634	P27pn_cytoBR	ERALPHApn_cytoBR	(CDK4pn_cytoBR,CDK2pn_cyto	input/output	2
protein	GC13P048303	HCC1954, SK-BR-3,	Breast	Human	Breast	cytoplasm	GO:0005737	PRBpn_cytoBR	6pn_cytoBR),(CDK4pn_cytoBR,CDK6pn_cytoBR,CDK2pn_cytoBR)		output	2

Figure 2 Standardized spreadsheet format

2) Graph-specific network characteristics: To evaluate each model from graph theory perspective, we proposed to measure benchmarks on number of nodes/edges, fan-in and fan-out cone, inner loop counting, and pathfinding analysis.

3) Executable model characteristics: Executable model analysis under different scenarios using both deterministic and stochastic simulations and generating heatmap visualization can provide insights into the transient behavior of the biological system, as well as the steady states that are reachable under these scenarios.

In the following subsections, we will discuss in detail the metrics we propose to evaluate both graph-based and executable model characteristics.

3.1 Number of nodes and edges

Computational models of signaling networks can vary in size, from small models of only a few nodes and edges, to large models of hundreds of nodes and edges. While the size of the model may not be a complicated metric, it is the first step in comparing models. The number of inputs and outputs of a model contribute heavily to model dynamics; therefore, this metric has also been utilized to describe the benchmark model. While many signaling networks are visualized with clear inputs (cytokines, exogenous chemicals, etc.) and outputs (genes, cell processes, etc.), this can be difficult for larger models. Additionally, different types of models may have different expected inputs and outputs, which can affect the behavior of the model. Identification of all outputs of the model can help identify potential observables of the model, as some outputs will be

cell processes such as “inflammation” or “cell cycle progression”, brought on by gene transcription.

3.2 Feedback loops

Stable biological systems, such as cells and organs, not only receive external stimulus, but also perceive their internal environment information. The information is used to make suitable adjustments through internal mechanisms [28], which is defined as a feedback system. Feedback loops are a common feature of signal transduction networks and play a critical role in regulating their behavior. This regulatory motif is capable of inducing complex behavior by connecting a downstream element to one of its regulators. While there are several types of feedback loops (negative, positive, dual negative and positive, etc.) all adapt the input of the model to its outputs. Feedback loops are common in signaling networks, such as the *p53* [29], *ERK*[30] and *Wnt* [31] pathways. If elements are connected as a directed cycle [32], then it is defined as a feedback loop. A simple *m* elements feedback loop contains *m* elements and *m* interconnected edges.

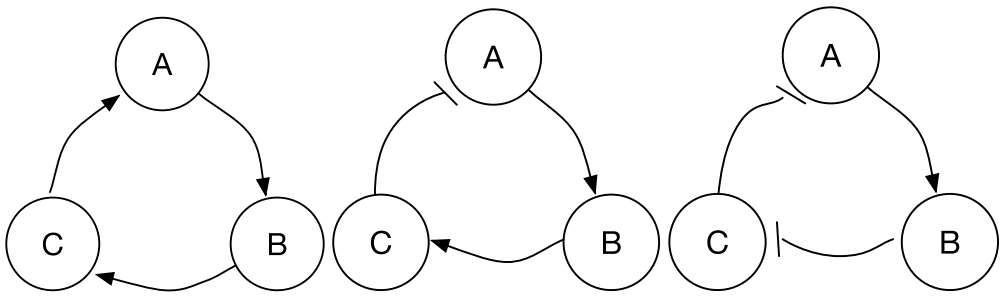

Figure 3 (left) Positive loop, (middle) Negative loop, (right) Dual negative feedback loop

Algorithm 4: Find all Positive and Negative feedback loops

```
for j in range(len(all_edges)):
    if len(all_edges[j] == 0):
        if len(self_loop_pos) < 1:
            pass
        else:
            posOne += 1
    if len(self_loop_neg) < 1:
            pass
    else:
            negOne += 1
    elif len(all_edges[j] == 1):
        if (pos_count % 2) == 0:
            pos_Two += 1
        elif (neg_count %2) == 1:
            neg_Two += 1
    else:
      for i in range (len(all_edges_sorted)):
            if (all_edge_sorted[i] % 2) == 1:
                  pos_mul += 1
            else:
                  neg_mul += 1
total_negLoop = negOne+ negTwo + neg_mul
total_posLoop = posOne + posTwo + pos_mul
```

Figure 4 Pseudocode for finding how many positive and negative feedback loops

3.3 Inner loop

An inner loop can be defined as a subset of a graph constituting a feedback loop, in which the nodes and edges of the subset are also connected in a directed cycle. Figure 5 is an example of a feedback loop containing several inner loops. The biggest feedback loop is *(A-B-C-D-A)*, which is a positive feedback loop, and has three inner loops. The algorithm described in Figure 6 enumerates the inner loops *(A-B-D-A)*, *(A-D-A)*, and *(A-C-D-A)*. While the largest feedback loop in Figure 5 is positive, all three inner loops are negative. Depending on the activity of the

elements *B*, *C*, and *D*, the overall effect element *A* has on its own activity may be inhibitory rather than activating. Additionally, incoherent feedback loops converge to different attractors than coherent loops, and show differences in robustness [33]. Identification of inner loops in biological signaling networks provides crucial information on complex regulatory behavior, considering that nested feedback loops in signaling pathways are exceedingly common [34].

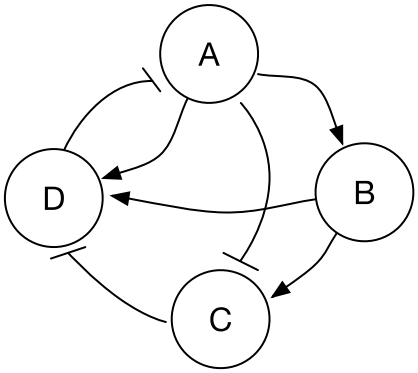

Figure 5 Toy example for an inner loop

Algorithm 2: Inner loop

```
1 count_subloop = 0
2 def sublist(lst1, lst2):
3     return set(lst1) <= set(lst2) # compare two sets
4 for i in range(len(feedback_loops))
5     for j in range(len(feedback_loops)):
6         if sublist(feedback_loops[i] = feedback_loops[j]) == True:
7             if feedback_loops[i] == feedback_loop[j]: # need to remove self-loop (redundancy)
8                 pass
9             else:
10                count_subloop += 1
11                print(str(feedback_loop[i] + ' is the inner loop of ' + str(feedback_loops[j])))
12 print('There are total '+ str(count_subloop) + ' inner loops in the model')
```

Figure 6 Pseudocode for the identification of inner loops

3.4 Pathfinding

Signaling networks models can be complex, with hundreds of genes and proteins connected by even more interactions. For this reason, finding paths between two elements in a model has been implemented previously [35]. We can use several different path metrics for describing the benchmark models- the shortest path, longest path, and all paths between a source and target. The first two metrics can identify differences in the timing of two different regulation mechanisms. By finding all paths, we can identify different regulatory mechanisms for a pair of elements.

Algorithm 3: All paths from input to output

```
multiPath = [ ]
large = max([len(Out), len(In)])
small = min([len(Out), len(In)])
if ((len(Out)>1) & (len(In)>1)) == 1:
    // Scenarios for multiple inputs and outputs
    for i in range(large):
        for j in range(small):
            path = graph.find_all_paths(Out[i],In(j))
            multipath.append(path)
print('There are total ' + str(len(multipath[0])) + ' difference paths between input and output')
else:
// Scenarios for single input and output
path = graph.find_all_paths(Out[i],In(j))
multiPath.append(path)
```

Figure 7 Pseudocode for the finding paths between all input and output

3.5 Fan-in and fan-out cone analysis

Albert's paper [36] discusses the concept of "centrality" in biological modeling, where researchers are usually interested in connectivity and functionality of each node. Similarly, for the benchmarking characteristics in this paper, identification of fan-in and fan-out cones will be discussed. First of all, the concept of fan-in and fan-out originates from digital electronic design. Fan-in is defined as the number of inputs a logic gate can handle [37]. For instance, the fan-in for the AND gate in Figure 8 is three. Similarly, the fan-out is defined as the number of gate inputs it can drive for a logic gate's output [37]. We are also interested in finding all predecessors and successors of an element. Thus, if we traverse all the predecessors of element A, we define it as the fan-in cone area of element A. On the other hand, all the successors of element A are defined as fan-out cone area. From Figure 9, it is clear that fan-in cone of element A contains three elements (B, C, D), and fan-out cone contains total five elements (E, F, G, H, I). Identification of the fan-in and fan-out cones for key elements in a signaling pathway can help illuminate differences between models. Specifically, this is a straightforward method to compare regulation of the same element between pathways.

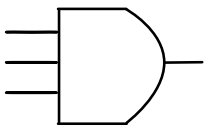

Figure 8 AND gate

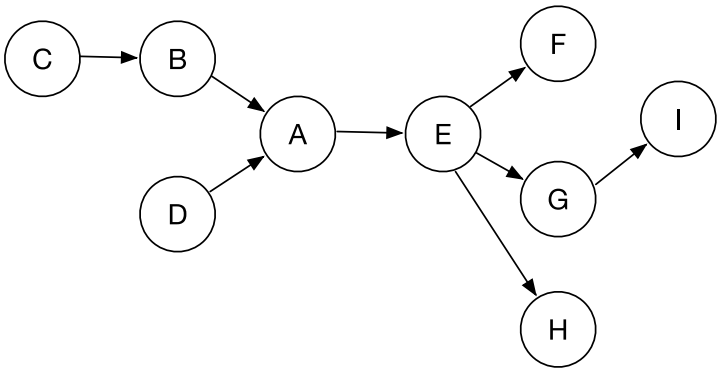

Figure 9 Fan-in and Fan-out cone

3.6 Scenario analysis

The final analysis of the benchmark models is the activity of model elements during a simulation. For all of our benchmark models, we use the DiSH simulator [26, 38], which is capable of both deterministic and stochastic simulations. With the right parameters, a simulation provides vital information on model behavior, and is a powerful standard for comparing models. For two models of the same disease or pathway, the activity of key elements is vital to comparing their accuracy. We can compare the steady state or end values of model elements to biological data, which further improves the characterization of our benchmark model.

Deterministic update scheme. Deterministic simulations execute all rules in a model in a pre-determined order. This order may be simultaneous or ranked-order. Model elements need much less time to reach a steady state than other simulation schemes [26]. We can use the number of steps needed to reach a steady state as a benchmark for model dynamics, in addition to model element end values. However, these end values may not be comprehensive.

Stochastic update scheme. Unlike the deterministic simulation scheme, the stochastic simulation scheme incorporates random choices when choosing which rules to execute. The rules for each update step can be chosen uniformly, or according to user-input probabilities. While this method takes more time, it is more indicative of a biological system. Different runs may have different element steady states due to the order of the selected update rules. Depending in the problem statement, stochastic or deterministic simulations may be more suitable.

Randomized initial values. For many situations, the inputs to a model may not be established. The starting activity of an element is highly context-specific, due to disease state, cell type, tissue, etc. To study the different states a model may reach, DiSH can use randomized initial values for model elements. Depending on the number of randomized initial values, this simulation scheme can traverse all possible model outcomes. We can use the possible model outcomes as a measure of how adaptive a benchmark model is.

3.6.1 Heatmap visualization

A heatmap of element activity over the course of the simulation can be treated as another useful benchmark. In this thesis, individual element values are represented with two colors (black or white). Black represents the OFF state, while white represents the ON state. Through heatmap visualization, we can better benchmark and visualize the model, where the final state and initial value can be observed in a cleaner fashion than the previous trace method. The heatmap benchmark is also able to succinctly portray element activity for all runs, not just the average over the course of the simulation. In Chapter 4.0, we will generate heatmap benchmarks for the breast cancer, angiogenesis, and Tcell model.

4.0 Benchmark specifics and discussions

The philosophy of designing different benchmarks are proposed in the last chapter. In this chapter, the detailed metrics are studied and presented for ABA plant model, LGL, breast cancer model, angiogenesis model, Tcell model, macrophage model, and PCC model. From the result, we can compare the differences between each model and make the conclusions as well. The hardware computation platform chosen for CABARET project was based on MacBook Pro 2017, macOS Catalina, Dual-Core Intel Core i5, and 16GB RAM.

4.1 ABA plant model

The biological system of this analysis is the signal transduction network that is activated when plants experience drought conditions. During drought, a type of hormone abscisic acid (ABA) in plants will be synthesized. Then, the stomatal closure will increase, and finally the internal water evaporation can be minimized to balance the water distribution. Previously, a comprehensive analysis has been published by Li et al. (2006) [39], which indicated the network includes oscillations. However, whether the oscillation can be sustained needs further work. The whole model contains 54 nodes that include proteins, ion channels and secondary messengers and a few conceptual nodes such as “depolarization” and “stomatal closure”.

In this paper by Albert et al. (2010) [11], a systematic study of long-term dynamic behavior was performed with a traditional synchronous method and three other asynchronous updating algorithms under both perturbed and unperturbed scenarios. Overall, the dynamic behavior of

stomatal closure is observed when a biologically significant node is perturbed, also due to the change of existing positive feedback loops or negative loops.

It is observed that fan-in and fan-out cone analysis in Figure 10 indicates 12 elements (dark blue bars) such as *ABA, GCR1, ABH1, AGB1, ERA1* have no elements in their fan-in cone area. In other words, those 12 elements are not regulated by any other elements, which means they should be close to the model input side. The dark green bars, such as *Closure, DAG, ROP10* indicate elements with a fan-out cone size of zero, it is not regulating any other elements, which means it should be acting as the model output. In Figure 11, the pie chart illustrates how many elements are in the feedback loops, and the distribution of the size of all feedback loops. In the ABA model, the most common feedback loop size contains four elements, which accounts for 36.36% of all loops. Whereas the loops that contain only two elements have the lowest frequency among all feedback loops. In Figure 12, the average in-degree and average out-degree indicate how many input or output edges directly connect to the element. Additionally, the average length for feedback loop, positive feedback loop, negative feedback loop could be provided as useful benchmarks, which the comparative result among them could evaluate the oscillation or homeostasis behavior.

Last but not least, researchers are usually interested in finding all paths between system input and output. In Figure 13, we can clearly see which path length exists the most frequently. Obviously for ABA model, the path between input and output that contains 13 nodes is the most frequent type among all paths. Also, the longest path traverses 18 nodes and it only exists a very few times. The benchmark analysis indicates that the observable *Closure* is dependent upon several types of input and is subject to complex regulatory behavior due to upstream feedback loops. However, this model also has several “unconnected” nodes, which have a fan-in and fan-out cone size of zero. This benchmark reveals a potential weakness of the ABA model.

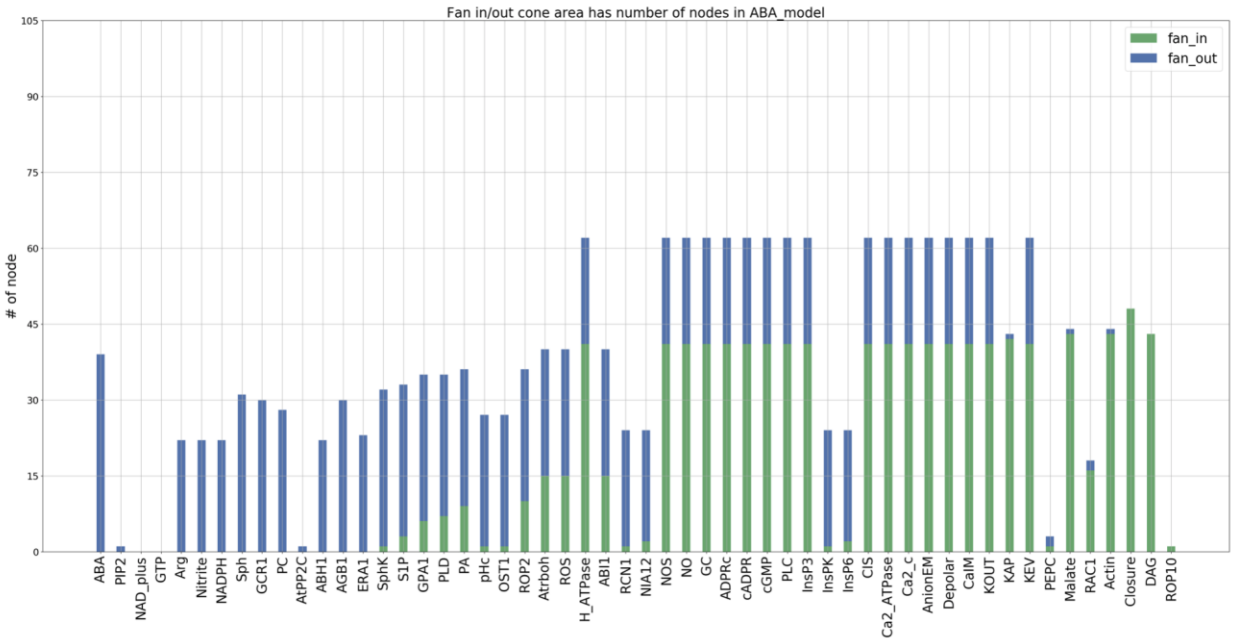

Figure 10 Fan-in and fan-out analysis for ABA model

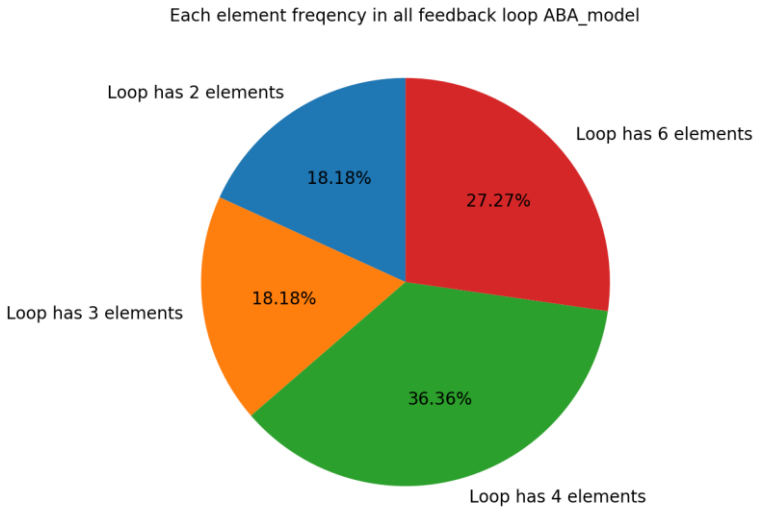

Figure 11 Feedback loop analysis for ABA model

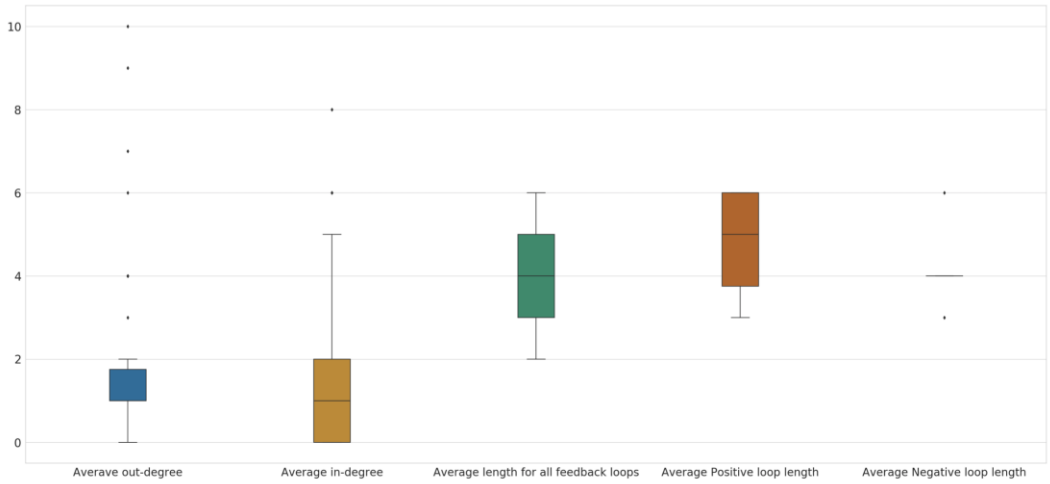

Figure 12 Average for other characteristics in ABA model

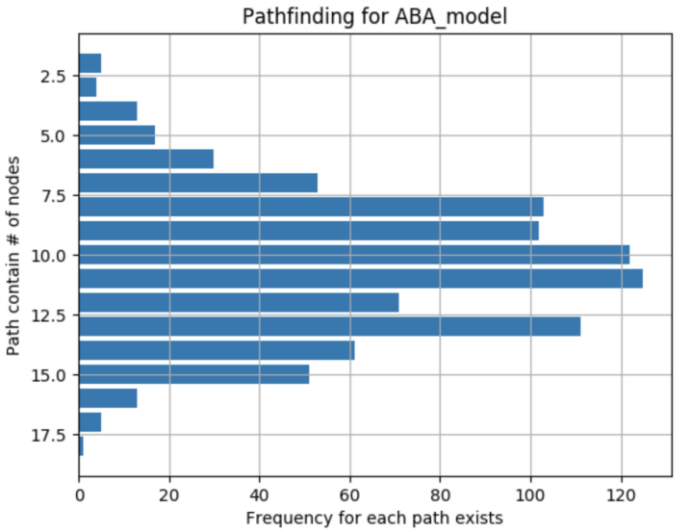

Figure 13 Information about pathfinding in ABA model

4.2 LGL model

T cell large granular lymphocyte (T-LGL) leukemia is characterized by abnormal clonal expansion of mature cytotoxic T lymphocytes (CTL). Thus, it can successfully escape activation-induced cell death (AICD) and remain competent for a long time. By extensive literature review and a Boolean model implementation, Albert et al. (2008) [12] found that the continuous expression of *IL-15* and *PDGF* is sufficient enough to reconstruct all the signaling abnormalities. Also, focusing on the model output *Apoptosis* in leukemia T-LGL, two nodes *SPHK1* and *NFkB* are verified experimentally. In other words, by inhibition of *SPHK1* and *NFkB*, apoptosis in T-LGL can be induced. Overall, the systematic re-construction of T-LGL model can help both biologists and clinical professionals, which provides them with the maximized available pathway information and possible key mediators in future therapeutic treatment.

From Figure 14, we can clearly observe elements like *Stimuli*, *IL15*, *PDGF*, *CD45*, *TAX*, *y1*(dummy node), *y2*(dummy node) have zero fan-in cone, which indicate they act as system inputs. Also, *Apoptosis*, *Cytoskeleton_signaling* and *Proliferation* contain zero fan-out cone element, which should be identified as system output. Additionally, we can clearly see large portion of homogeneous fan-in and fan-out cone elements, such as *LCK*, *ZAP70*, *GRB2*, *PLCG1*, etc., which is due to the large number of cycles in the model.

In Figure 15, there are various types of length for feedback loops, the feedback loop that contains 20 elements is the most frequent type. For simplicity, the rest of 19 different feedback loops are grouped together. From the average characteristics in Figure 16, LGL model has approximately the same average length for all positive and negative feedback loops which can be further studied. Finally, in Figure 17, the normal distribution shape informs us the length between

input and output which reaches the peak at 25 elements. Also, the shortest path needs to traverse at least 9 nodes from the input to the output.

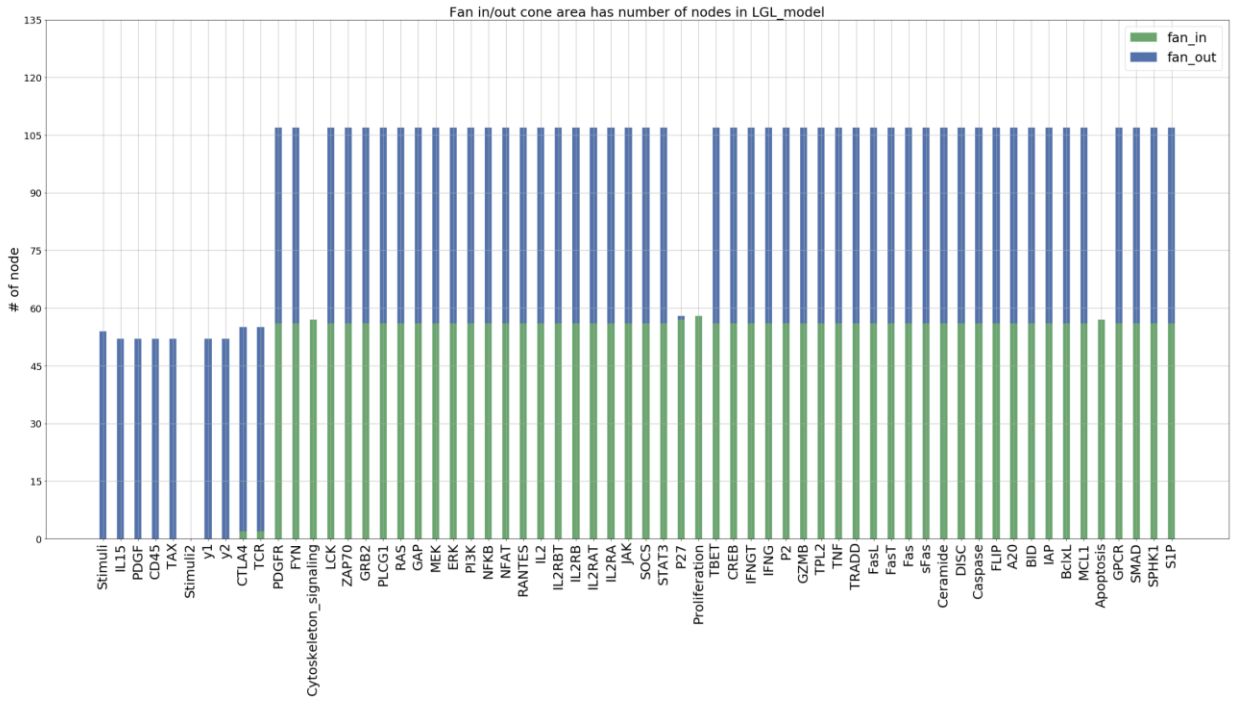

Figure 14 Fan-in and fan-out analysis for LGL model

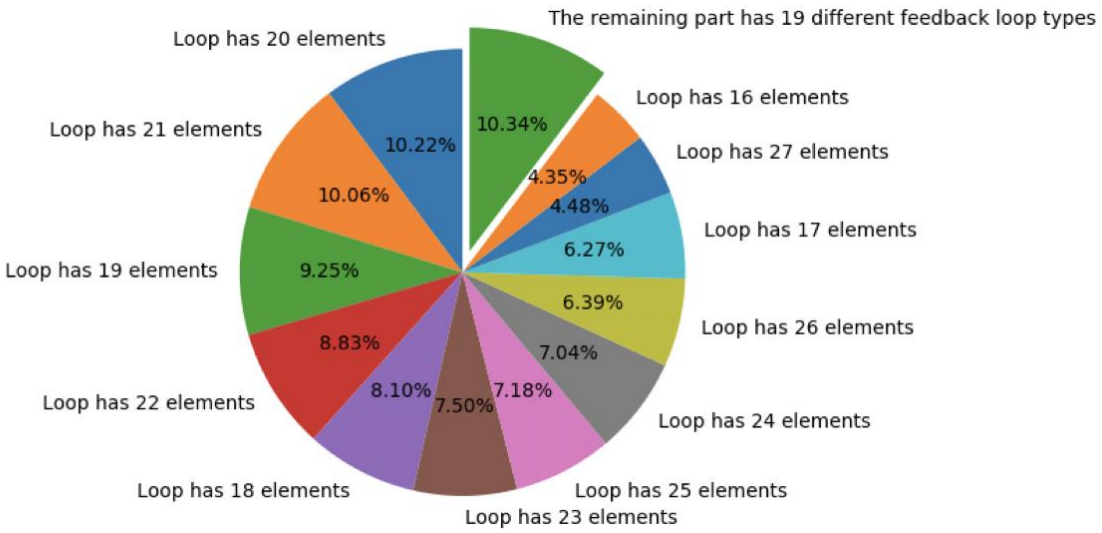

Figure 15 Feedback loop analysis for LGL model

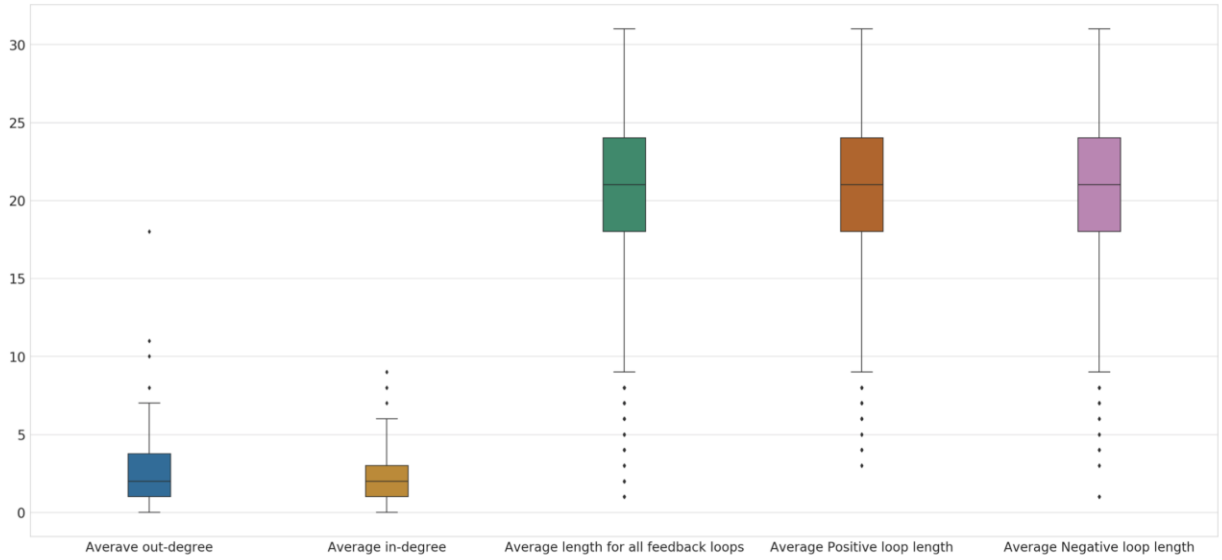

Figure 16 Average for other characteristics in LGL model

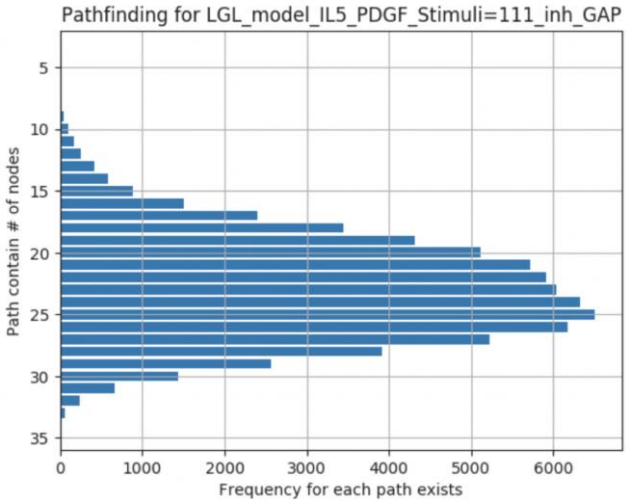

Figure 17 Information about pathfinding in LGL model

4.3 Breast cancer model

In breast cancer, overexpression of protein HER2 is an adverse prognostic marker, which occurs in almost 30% of patients. For clinical intervention, HER2 is targeted by monoclonal antibody trastuzumab. In reality, *de novo* resistance to this antibody remains a serious problem. To solve this issue, precise modeling and knockdown experiments are implemented to find possible novel therapeutic targets for trastuzumab resistant breast cancer [13].

Extracting model for HER2 receptor-regulated G1/S transition from various publications, which the whole model contains 20 protein elements from input *EGF* to output *pRB* (phosphorylated tumor suppressor retinoblastoma protein). Biologically, hyperphosphorylation of *pRB* will lead to the release of E2F transcription factor, initiating the transcription of essential genes for DNA replication. In other words, in normal and tumor cells, *pRB* will toggle between the active (hypophosphorylated) state and the inactive (hyperphosphorylated) state. Correspondingly, the phosphorylation and subsequent inactivation of *pRB* stands for a marker for cell growth (proliferation).

From Figure 18, we can see that *EGF* does not have any fan-in cone, which verifies the original paper's description [13], that *EGF* is the input for the breast cancer model. Similarly, and naturally, *pRB* with zero fan-out cone verifies the original output statement as well. For the feedback analysis in Figure 19, over 57% feedback loops contain more than 2 elements. Another interesting point can be observed from Figure 20, that there is no negative feedback loop in the breast cancer model. Lastly, the longest length between input and output traverses 13 nodes and it exists 180 times. On the contrary, the shortest path contains only six nodes and it shows up around 10 times. From the knockdown experiment perspective, if we can know which path contains

targeted node *CMYC*, which happens to be one of the shortest lengths, we might able to speed up the inhibition analysis.

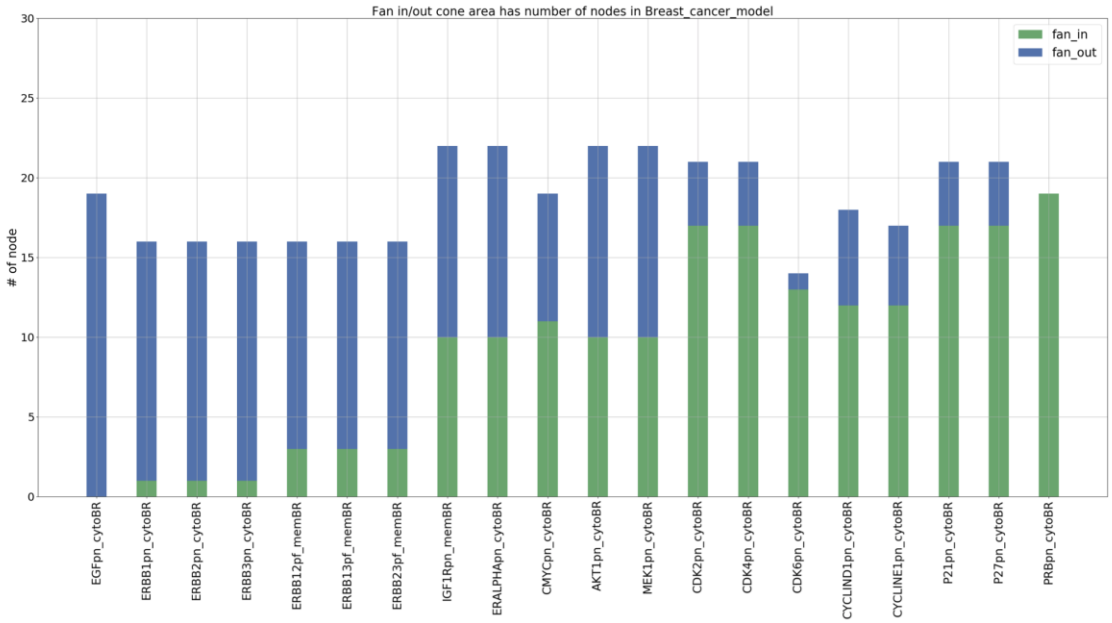

Figure 18 Fan-in and fan-out analysis for breast cancer model

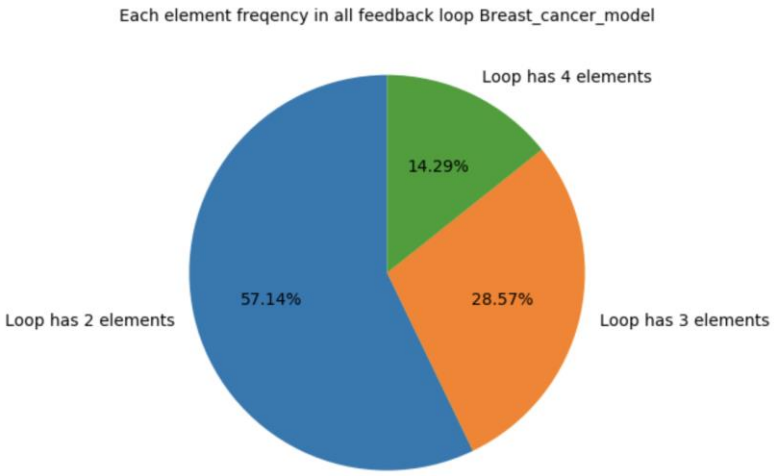

Figure 19 Feedback loop analysis for breast cancer model

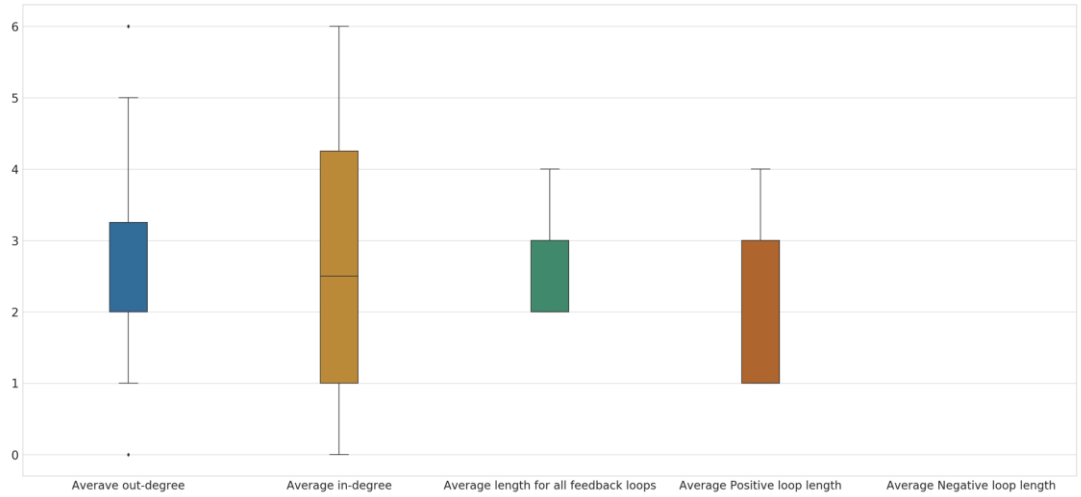

Figure 20 Average for other characteristics in breast cancer model

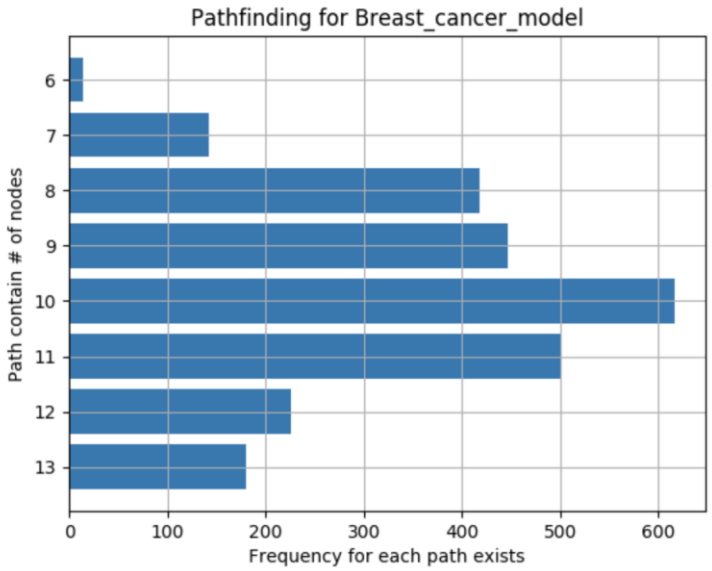

Figure 21 Information about pathfinding in breast cancer model

Normally, there will be many experimental scenarios in one model. However, from thesis simplicity perspective, I will only demonstrate heatmap results for one key knockdown. In the original paper [13], the author found out by knocking down element *CMYC*, it will significantly decrease output *pRB* compared with the unperturbed case. The two plots in Figure 22 and Figure 23 illustrate the unperturbed case, where *CMYC* will not be knocked down, and it will remain in high state. Correspondingly, output *pRB* will reach high state in the end. Not only steady state can be used as benchmark metrics, also initial values can be easily observed and confirmed, such as how initial values for *CMYC* and *pRB* are all randomly chosen shown in Figure 22 and Figure 23.

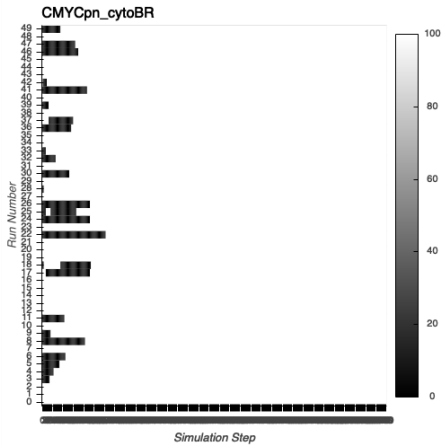

Figure 22 Unperturbed CMYC

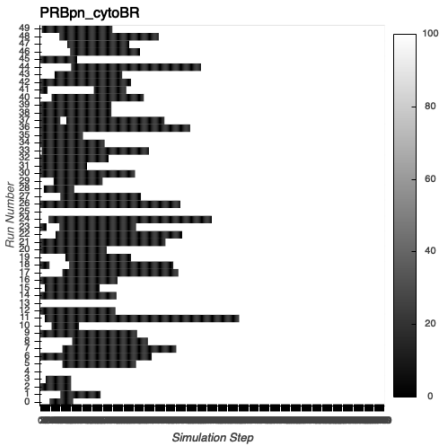

Figure 23 Unperturbed pRB

On the contrary, if *CMYC* knockdown is implemented, the *pRB* will show the opposite final steady state, which can be seen in Figure 24 and Figure 25. In other words, the heatmap results verify the paper's proposal, which *CMYC* could be used as knockdown node in future targeted therapy.

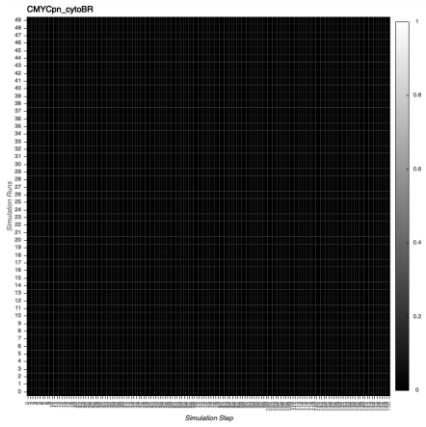

Figure 24 Perturbed CMYC

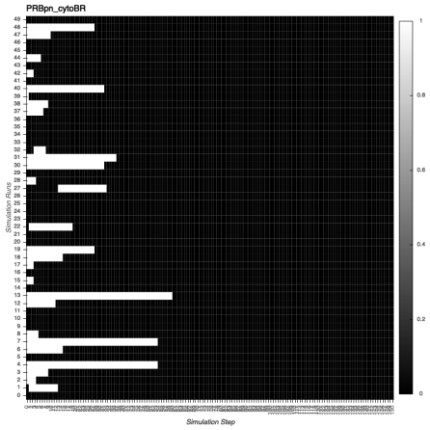

Figure 25 Perturbed pRB

4.4 Angiogenesis model

Tumors rely on a steady supply of oxygen and nutrients to maintain bulk. Cancer cells may induce blood vessel growth and development (angiogenesis) through emission of certain biochemical signals, leading to increased cancer invasion and metastasis. Without an increased blood supply, tumors are unable to proliferate at the same rate, which makes tumor-induced angiogenesis an attractive target for cancer treatments. Understanding precisely how endothelial cells synthesize multiple biochemical signals might help us to create novel therapeutic strategies.

Based on the experimental data from published literature, the model network contains three inputs: *VEGF-RTK, integrin, cadherin*. *Proliferation, apoptosis* and *motility* are the model outputs. In this paper [14], the author tested all possible input combinations and found out the output configurations are independent of any internal node initial value. Also, a few other scenarios like inhibition analysis and feedback analysis (between two key crosstalk elements *Rac1* and *RhoA*) are discussed as well. Through the experiments, we can discover a possible therapeutic approach

to control angiogenesis, which is based on identifying key element pathways. Furthermore, we can apply the similar method for future cancer study.

From Fan-in and Fan-out analysis in Figure 26, we can see *cadherin*, *VEGF_RTK* and *ITG_integrin* are the system inputs, where *VEGF_RTK* and *ITG_integrin* can cover more elements than cadherin. On the contrary, three outputs *Proliferation*, *Apoptosis* and *Motility* are reported, where *Proliferation* has higher capacity (can cover 19 nodes) than the rest two output nodes. If we compare the fan-in and fan-out cone analysis with previous LGL model, it is easily observed that the angiogenesis model has a considerably more heterogeneous fan-in and fan-out cone distribution for each element, which is mainly due to a no feedback system. Similarly, we have an empty box plot in Figure 27 for average feedback loop length because no cycle exists. As for Figure 28, we can clearly see the shortest path between input and output contains three elements and it only shows up one time. Additionally, the longest path has 11 nodes and traverses only one time as well.

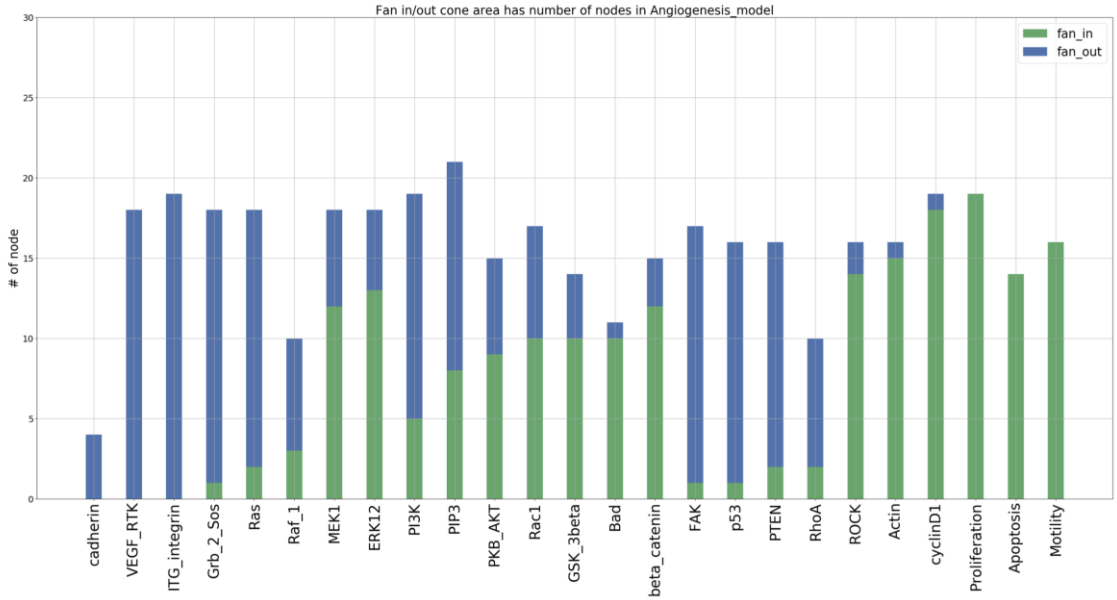

Figure 26 Fan-in and fan-out analysis for angiogenesis model

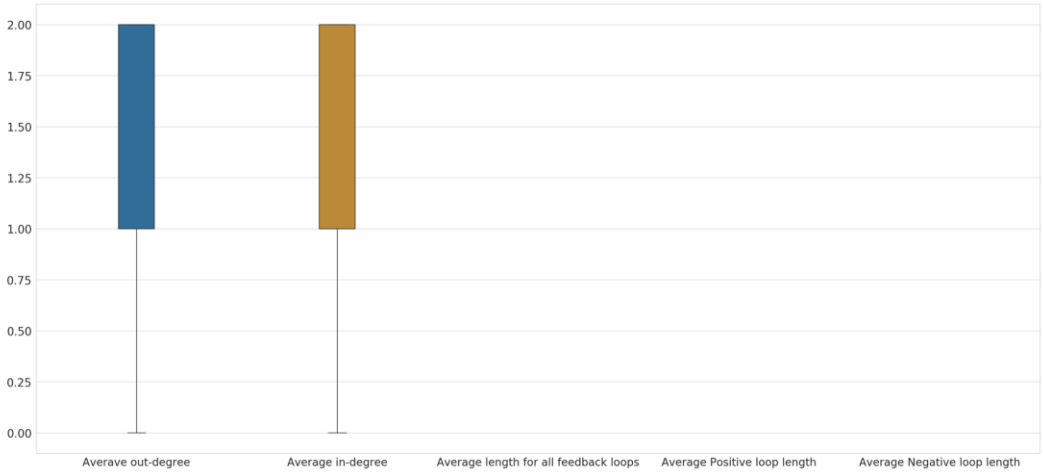

Figure 27 Average for other characteristics for angiogenesis model

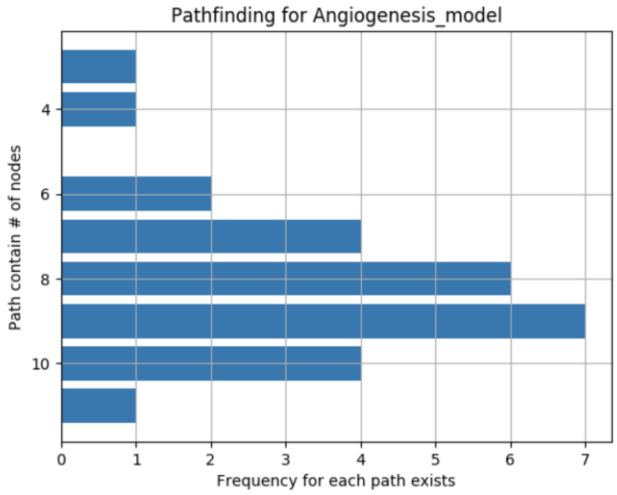

Figure 28 Information about pathfinding in angiogenesis model

There are two key scenarios in angiogenesis model, where the first scenario focuses on input-output map that uniquely characterize cellular response to output stimuli. The iterative

simulations from DiSH generate heatmaps for total 16 sets of input-output pairs. For simplicity, I only demonstrate one set of them. *Cadherin*, *VEGF_RTK*, *ITG_integrin*, *Rac (De)activation* are the four inputs. When input combination is 0111 (Rac is activated), the output cell phenotype should contain both proliferation and motility which are all turned ON in Figure 32 and Figure 33. This finding contradicts the widely believed "go and grow" hypothesis, which means cell motility and proliferation cannot happen simultaneously. However, Figure 32 and Figure 33 not only confirm the original paper's result [14], also, the time in x-axis indicates when the proliferation and motility will happen at the same time, or which one happens earlier. While the presence of all inputs increases both proliferation and motility, angiogenesis is turned OFF at the end of the simulation in Figure 34.

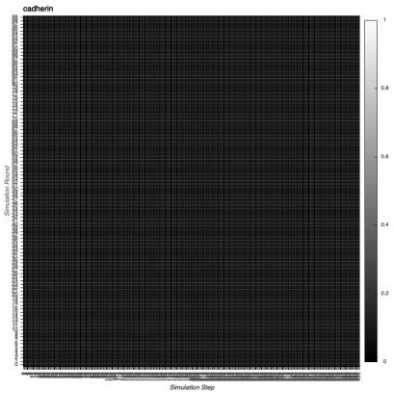

Figure 29 Cadherin

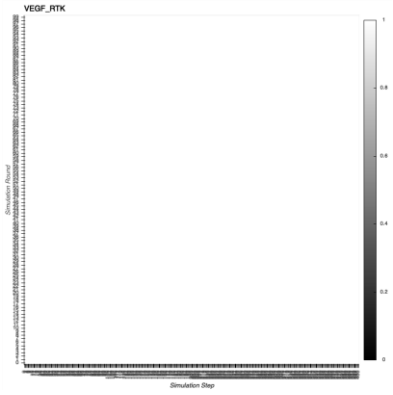

Figure 30 VEGF_RTK

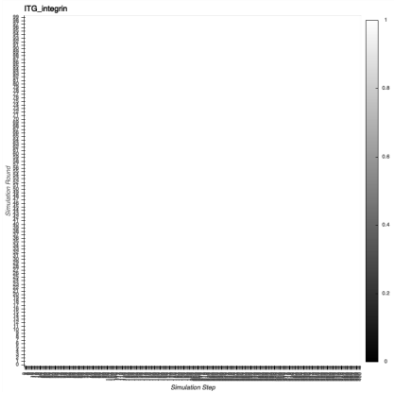

Figure 31 ITG_integrin

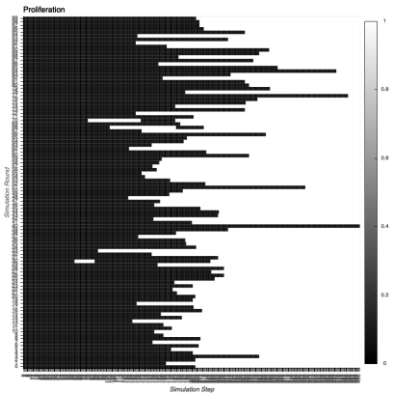

Figure 32 Proliferation

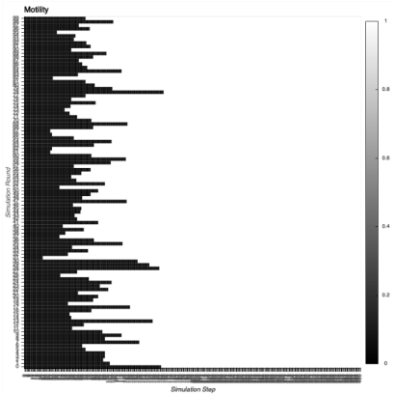

Figure 33 Motility

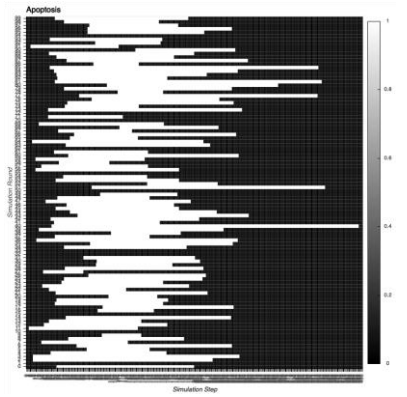

Figure 34 Angiogenesis

The second scenario concentrates on inhibition analysis, where the author is trying to find out whether targeted inhibition of specific signaling molecules will induce expected anti-angiogenesis effects. The inhibition experimental results indicate six elements *Grb-2/Sos*, *Ras*, *PIP3*, *Rac*, *GSK-3β* or *FAK*, they can all lead to apoptosis behavior regardless of any input configurations. For simplicity, *PIP3* is selected for demonstrating inhibition analysis under input combination 0111. The benchmarked outputs are clear, *Proliferation* and *Motility* are all turned OFF, and apoptosis will be induced in Figure 41.

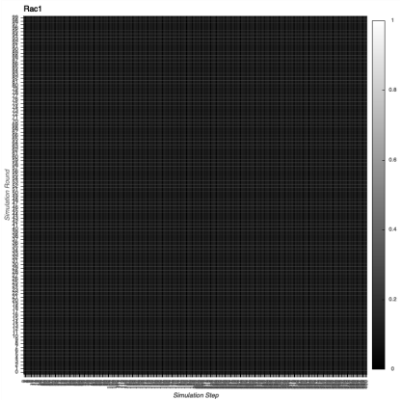

Figure 35 Rac

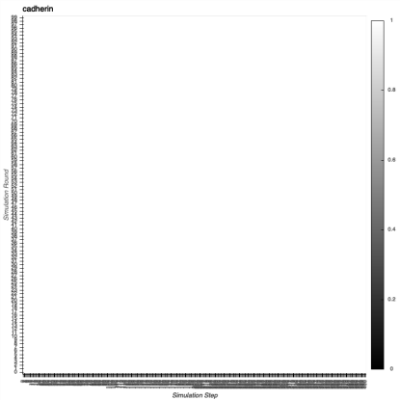

Figure 36 Cadherin

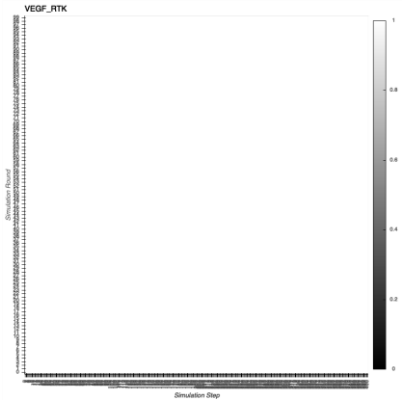

Figure 37 VEGF_RTK

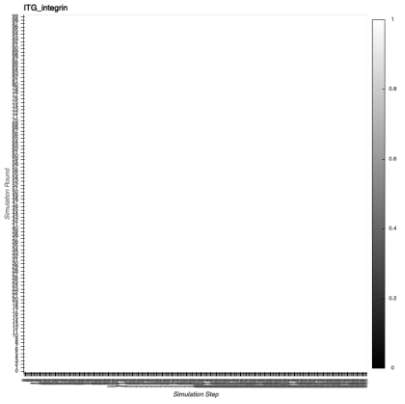

Figure 38 ITG_integrin

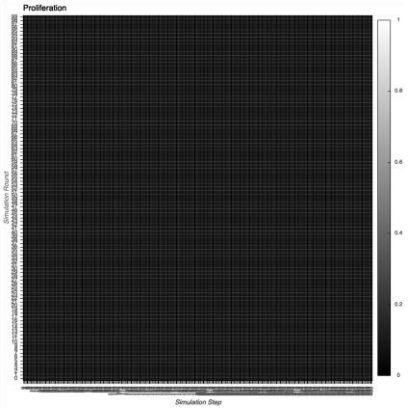

Figure 39 Proliferation

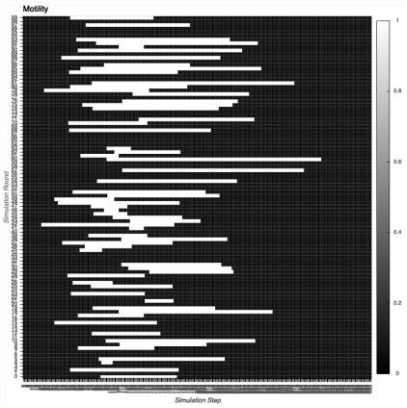

Figure 40 Motility

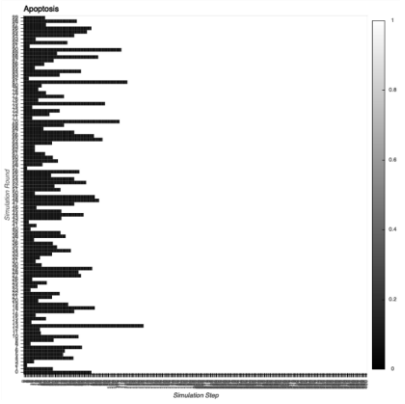

Figure 41 Angiogenesis

4.5 Tcell model

$CD4^+$ T cells has two sub-types, T_{helper} cell that can facilitate the immune response, and the $T_{regulator}$ cell which can suppress the immune response. In other words, patients diagnosed with increased numbers of $T_{regulator}$ might show greater suppressive immune system than healthy individuals. Also, $T_{regulator}$ cells are characterized by *Foxp3* (transcription factor forkhead box P3), which is a special pattern of cytokine production, and immunosuppressive function [15].

With various simulation in different delay and doses scenarios, such as the removal of antigen at different time points during experiment [15], then we can observe three types of cells: $CD25^{-}Foxp3^{-}$, $CD25^{+}Foxp3^{-}$, $CD25^{+}Foxp3^{+}$, which represents non-activated cells, $\mathrm{T_{helper}}$ cells, and $\mathrm{T_{regulator}}$ cells. Accordingly, the author found out the differential activation of TCR signaling pathways should be better modeled as various durations, rather than as signal intensity.

As we can see in Figure 42, six elements *TCR_low*, *TCR_high*, *CD28*, *IL2R_beta*, *IL2R_gamma* and *TGF_beta* contain zero fan-in cone, which indicates they should be close to the system input. On the contrary, *PI3K* is closed to the output due to zero fan-out cone. For the feedback loop length analysis in Figure 43, the top three popular loops contain 14, 18 and 12 elements, which can take more than 33% of the total feedback loop type. As for the nature of the feedback shown in Figure 44, Tcell model has a longer average length in negative loop (18 nodes) than positive loop (14 nodes). Lastly, for path analysis, Tcell model has the shortest path of 3 elements and it only exists a few times. On the other side, the longest path of 29 nodes shows up approximately 90 times. Moreover, the most frequent paths between input and output have 23 nodes.

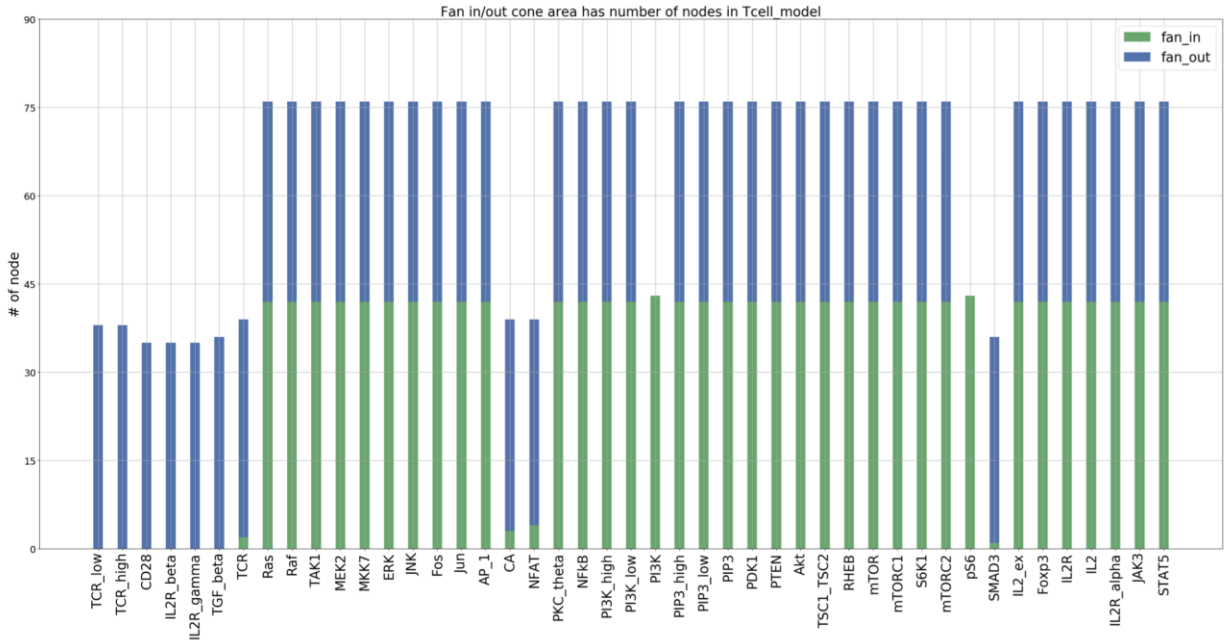

Figure 42 Fan-in and fan-out analysis for Tcell model

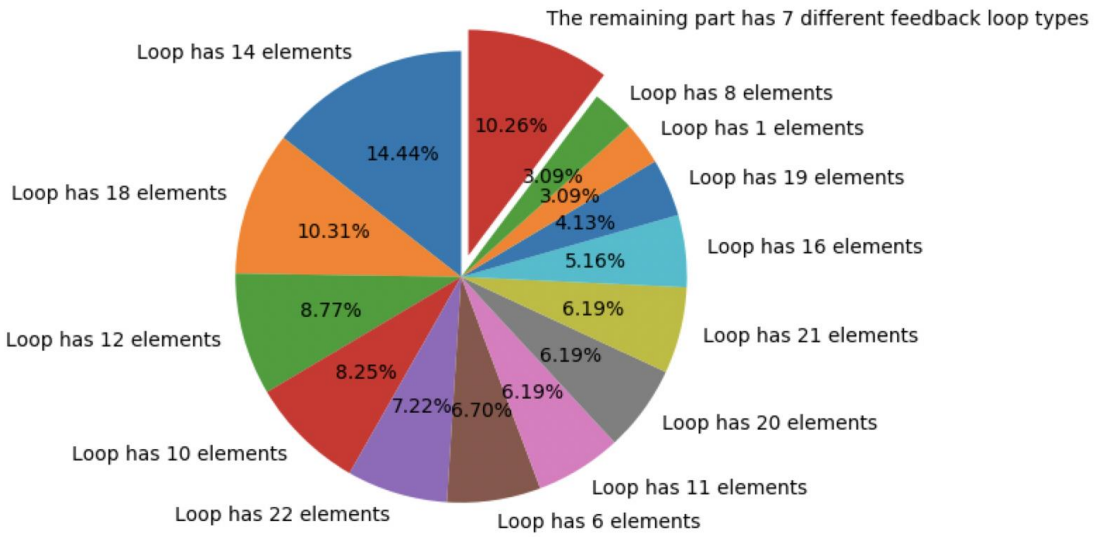

Figure 43 Feedback loop analysis for Tcell model

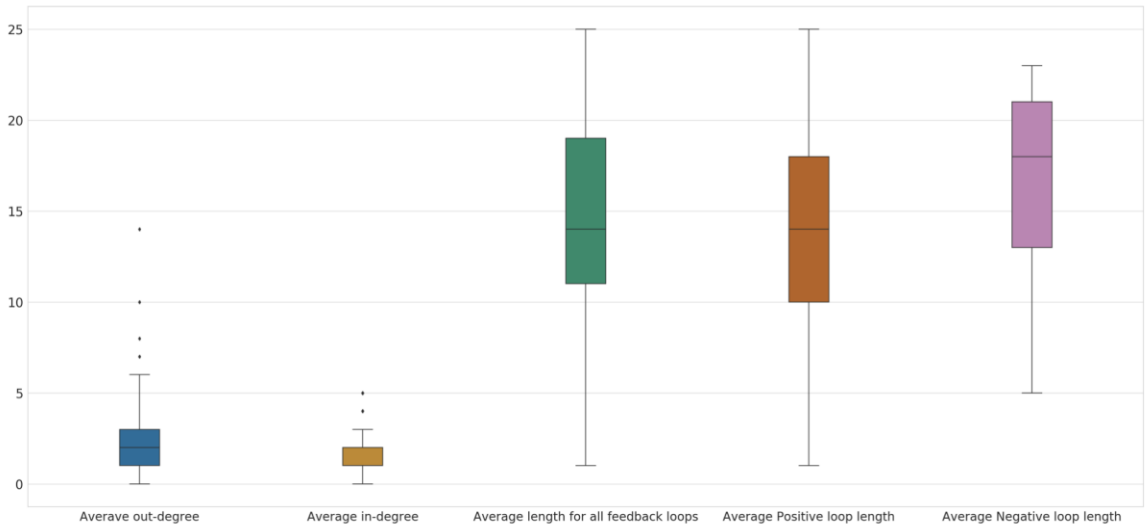

Figure 44 Average for other characteristics in Tcell model

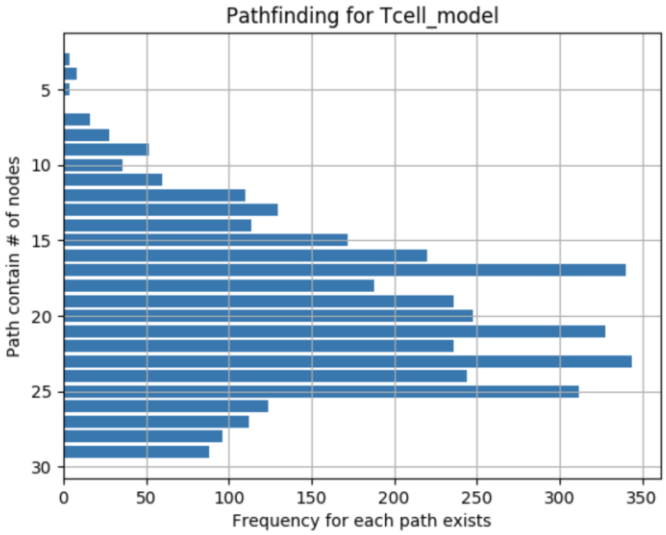

Figure 45 Information about pathfinding in Tcell model

In this thesis, we test the model under low or high concentration of antigen (TCR). Some molecules, such as *PTEN*, *Akt*, *PDK1*, *pS6*, *RHEB*, *S6K1*, *mTORC*, and *mTORC1*, have activity

that is independent of *TCR* concentration. However, *Foxp3* and *IL2* are dependent upon TCR concentration. For simplicity reason, the heatmaps shown in Figure 46 – Figure 53 can clearly illustrate the output differences between *Foxp3* and *IL2*. *Foxp3* is activated by low antigen dose, while *IL2* is activated by high antigen dose. In the presence of a high dose of *TCR*, T cells are susceptible to differentiating into $T_{regulator}$ cells.

High dose scenario:

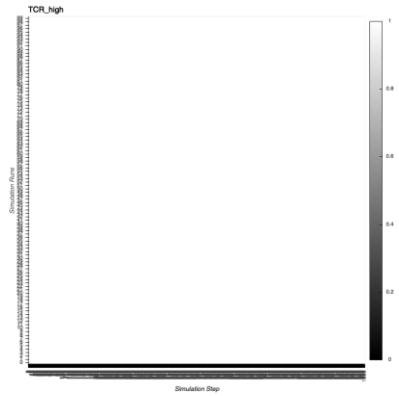

Figure 46 TCR_high = 1

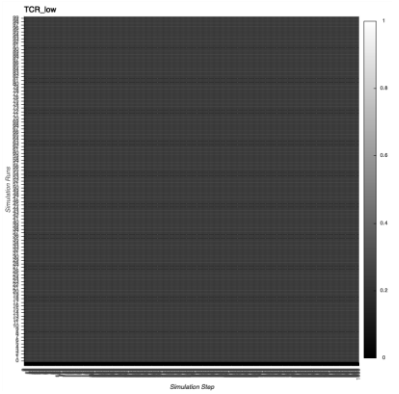

Figure 47 TCR_low = 0

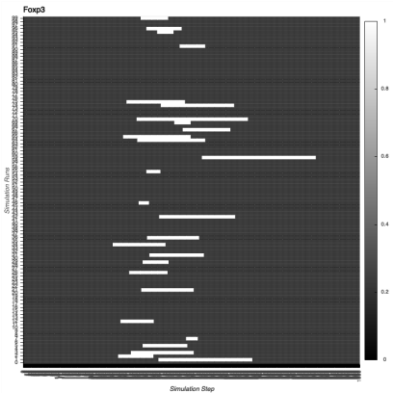

Figure 48 Foxp3 = 0

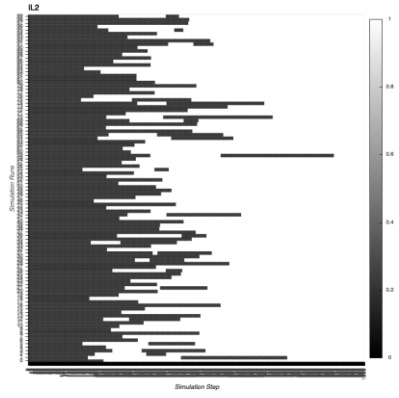

Figure 49 IL2 = 1

Low dose scenario:

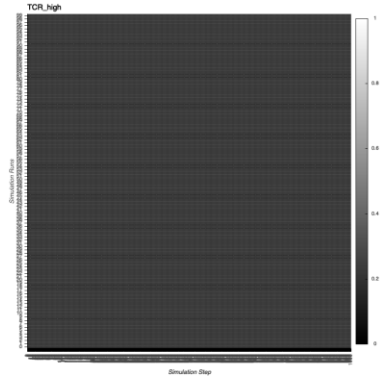

Figure 50 TCR_high = 0

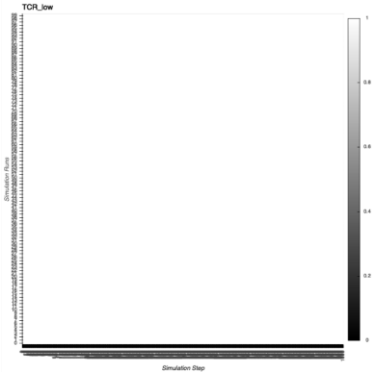

Figure 51 TCR_low = 1

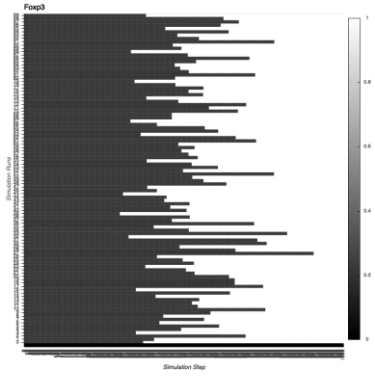

Figure 52 Foxp3 = 1

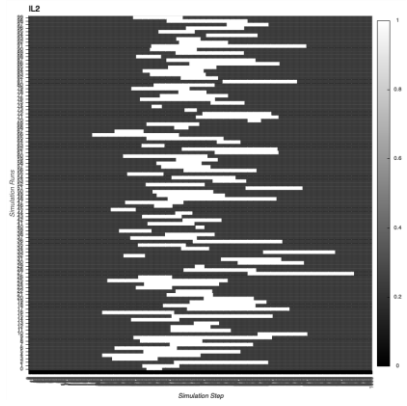

Figure 53 IL2 = 0

4.6 Macrophage model

Macrophages are ubiquitous throughout the body. While most macrophages reside in tissue, they can also mature from bone marrow monocytes circulating in the blood. Tissue-resident macrophages Macrophages are an incredibly diverse cell type, with almost as many tissue-specific variations as types of tissue in the body. Under normal conditions, macrophages are considered

part of the innate immune system, but they also provide key janitorial functions for the body by removing waste and clearing cellular debris. Under abnormal conditions, they have been shown to play a significant role in various cancers [40], arthritis[41], and diabetes[42], among others. Commonly, macrophages are broadly defined based aon which of these functions they serve: M1, considered classically activated macrophages, are triggered after bacterial infections and can elicit a powerful inflammatory response; M2, or alternatively activated macrophages, in contrast, are associated with an antiinflammatory response and wound healing. This simplistic paradigm is frequently criticized for its failure to account for the rich diversity seen in context-specific macrophage activation (M2a, M2b, M2c, Tumor Associated Macrophages or TAMs, etc.)[43-45].

Due to their prevalence in many different diseases, they are frequently the subject of models, or heavily featured in others [46-53]. In our model, each type of molecule is refered to as a model element which can represent proteins, genes, and mRNAs. Each element is given: a discrete variable that represents its molecular concentration or activity at a given time; and an update function that combines the influence of all the regulators on the element and determines the element's next value. Simulation of the model takes each element's initial values and uses the update function to calculate its next value. Using out model, we were able to simulate the process of M1 activation using changes in ligand concentrations to elicit the phenotypes of each activation state. In addition, this model has undergone a preliminary round of validation and has been shown to accurately simulate M1 activation. This means that our current understanding of the differentiation pathway is sufficient to build functional differentiation models.

As we can see in fan-in and fan-out cone analysis in Figure 54, 13 elements do not have any regulators, which are *LPS*, *GlycosEndProdcf*, *IL4*, *IL13*, *TGFB1*, *CSF1R*, *TLR4*, *INAR1*, *INGR2*, *IL10R1*, *IL13RA1*, *IL4RGC*, *TGFR1*. In other words, they should act as system inputs.

Similarly, elements with zero fan-out cone are recognized as system output, which are *IL6*, *IL10*, *IL12*, *ARG*, *CD36*, *CD48*, *TNFA*. From Figure 55, interestingly, over 50% feedback loops only have two elements. Also, 18.5% comes from a single element loop, which means self-regulating type happens frequently as well in Macrophage model. Moreover, in Figure 57, in average length comparison, the Macrophage model has longer average length in positive loop than negative loop.

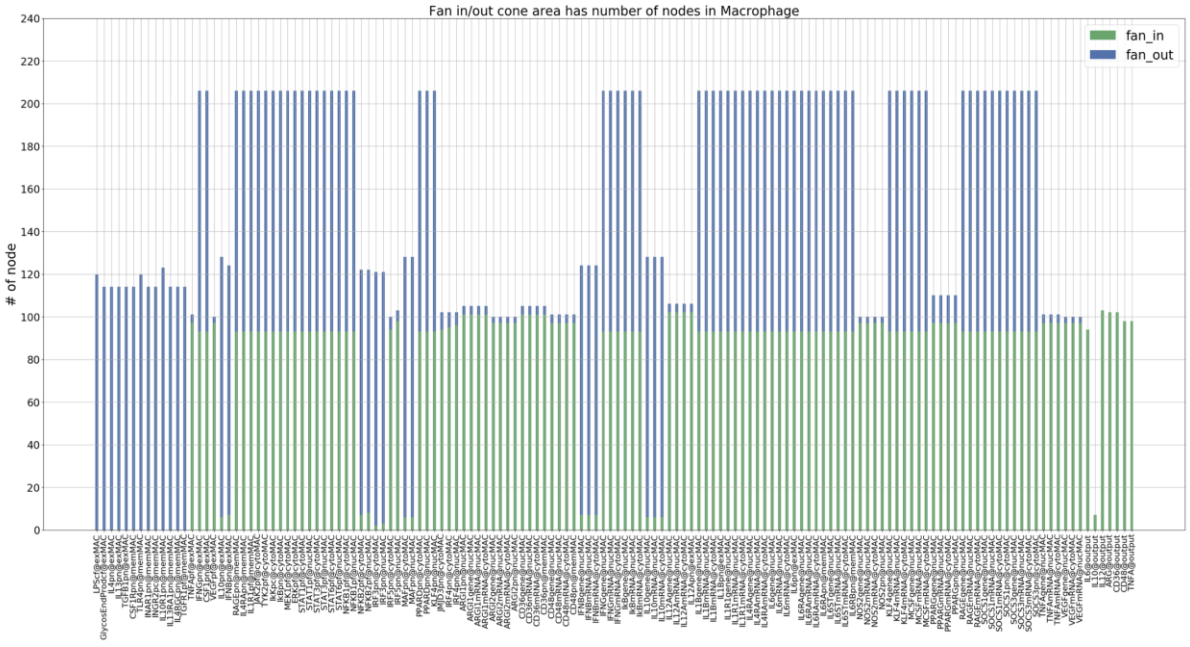

Figure 54 Fan-in and fan-out analysis for Macrophage model

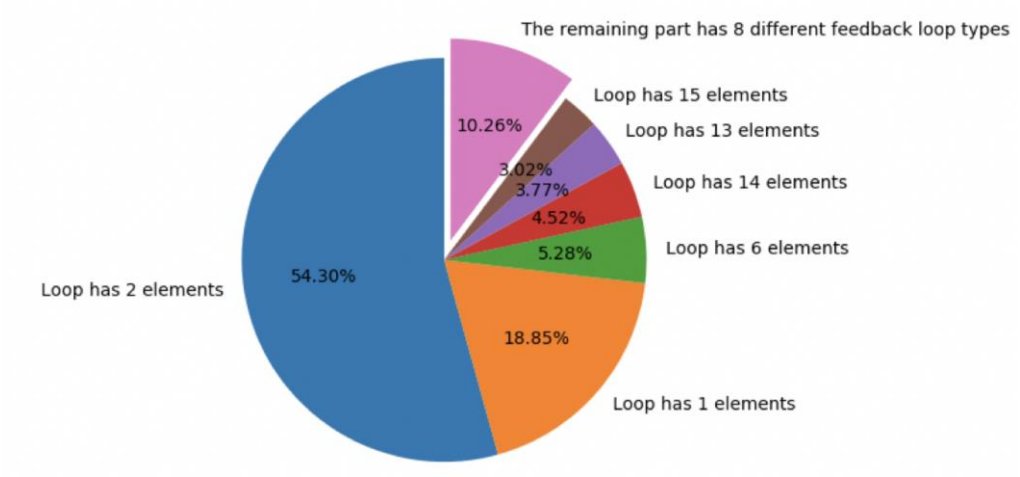

Figure 55 Feedback loop analysis for Macrophage model

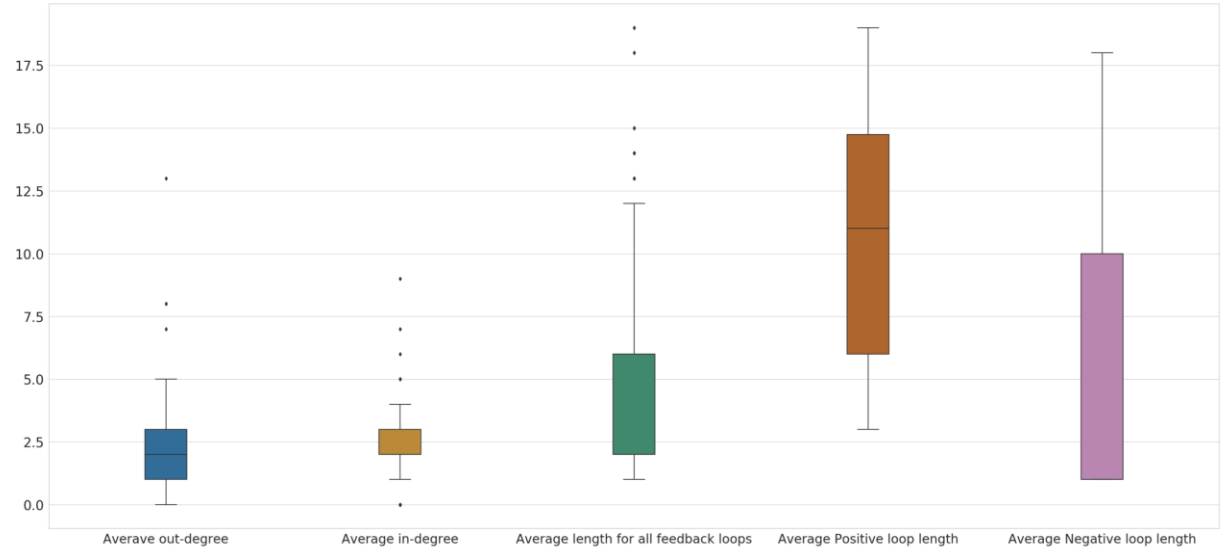

Figure 56 Average for other characteristics in Macrophage model

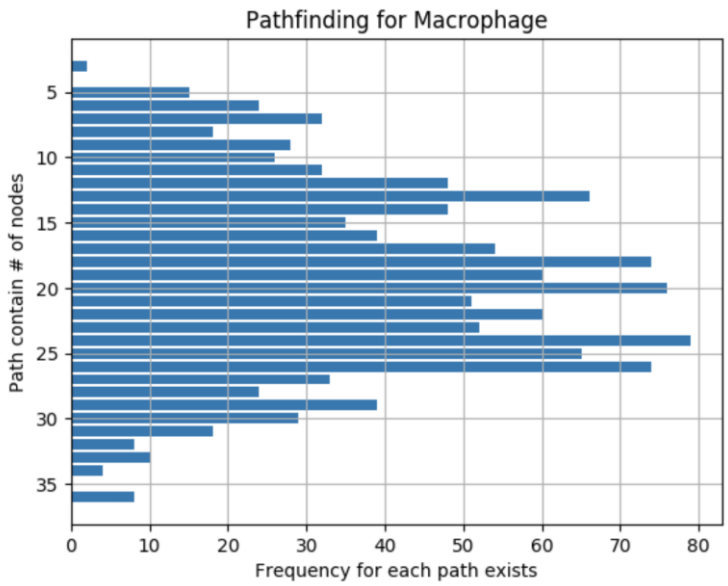

Figure 57 Information about pathfinding in Macrophage model

4.7 PCC model

Pancreatic cancer has very poor survival rate and the estimation shows that approximately 44,330 people will die of pancreatic cancer in 2018 [1 Boveri 2008]. Due to the early activation of KRas mutation followed by TP53 and CDN2A inactivating mutation in most tumor cells [Zeitouni, Pylayeva-Gupta et al. 2016, Raphael, Hruban et al. 2017]. Hence, targeting KRas mutation and restore tumor suppressive function becomes a daunting task. With various biological modeling approaches, it helps biomedical community to better understand the complexity of disease.

This new modeling approach is able to capture the dynamics features, timing of communication between cancer cells and their microenvironment, which the method contains parameterization and stochastic simulations [17]. Accordingly, the simulation results indicate the

influential signaling pathways and some key feedback and feedforward loops, which can be applied for future therapeutic study. Finally, with this precise modeling and signaling reconstruction, it will benefit more personalized cancer treatment and improve the quality of life.

For a large model like PCC, the fan-in and fan-out cone plot can give the audience a bigger picture. For example, *CCR2*, *CSF1R*, *CSF1*, *DUSP13* have zero fan-in cone, which indicates they are close to the system input. Whereas, elements like *APC*, *CDK4* and *AAHPROLIFERATION* have zero fan-out cone, meaning they are close to the output. Although *CDK4* and *AAHPROLIFERATION* both have zero fan-out cone, *AAHPROLIFERATION* contains much higher capacity in fan-in cone than *CDK4*, biologists might use this benchmark metric for the perturbed or knockdown study. As for the feedback path analysis in Figure 59, obviously, the PCC model has a highly uniformed feedback loop length distribution, which the top three feedback loops contain 44, 47, and 48 elements. Additionally, in PCC model, the average length for negative loop is a little bit higher than the positive loop which can be further investigated. Also, the reader might find out there is no pathfinding result for PCC model, because of 57 million paths between input and output, which exceed the computation power of our hardware platform.

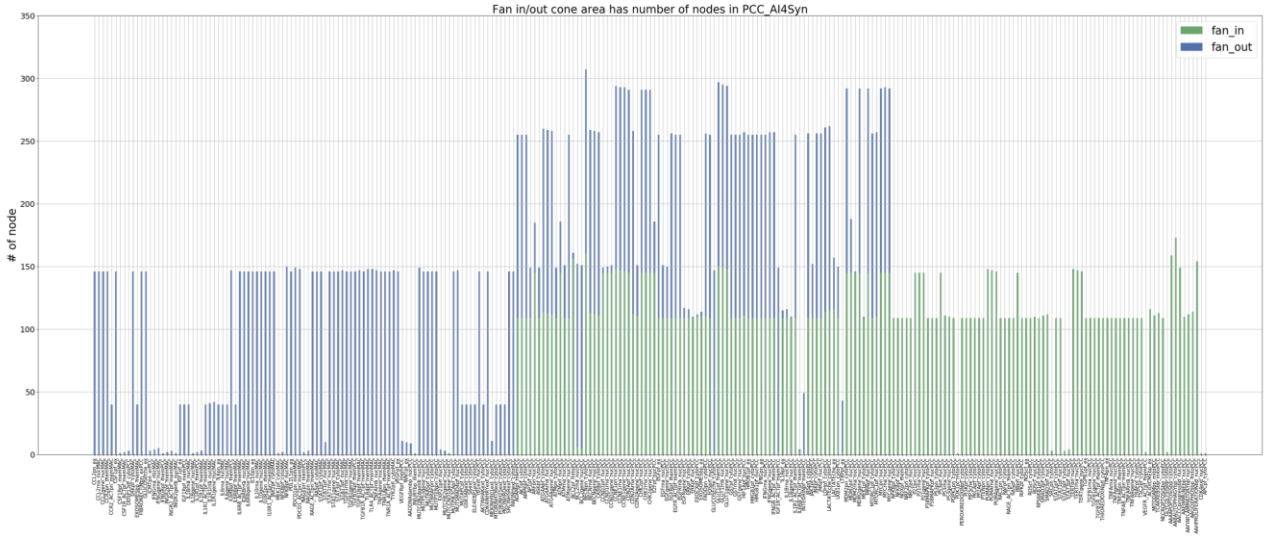

Figure 58 Fan-in and fan-out analysis for PCC model

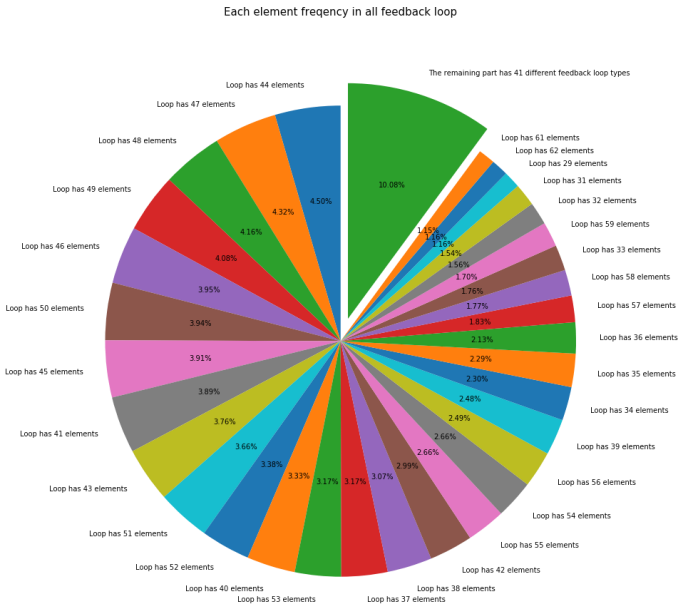

Figure 59 Feedback loop analysis for PCC model

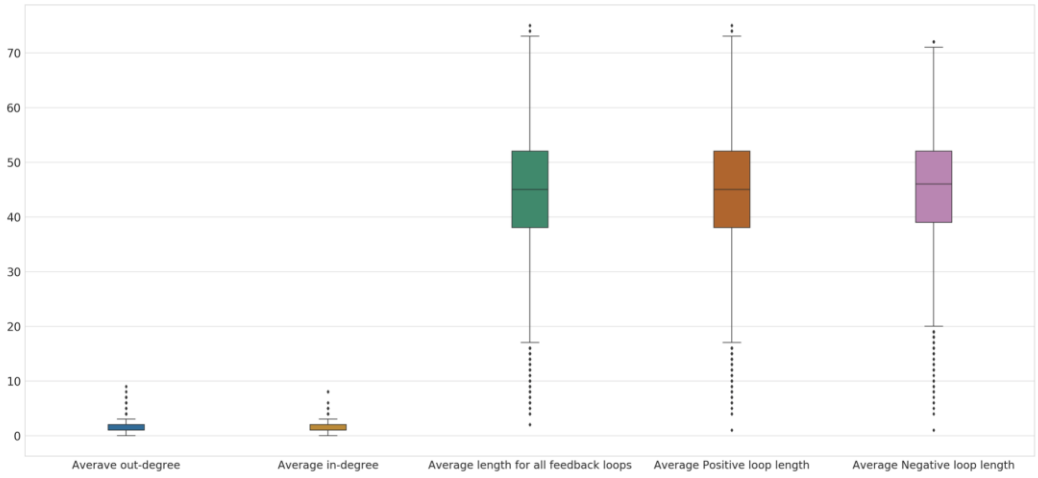

Figure 60 Average for other characteristics in PCC model

4.8 Summary for all benchmarked models

The benchmarking metrics for all models can be observed in Table 1, which the table includes a small model like breast cancer, a medium size model like Tcell model, to a complex model like PCC. Within the Table 1, the number of total feedback loops is not linearly increasing with a larger size of the model. Whereas, inner loop counting represents the membership from one loop to the other, the larger the number meaning we have a particular feedback loop is the subset of other loops many times. In the future work, more studies can be done for analyzing the redundancy problem of the inner loop. As for pathfinding metric, the total paths between input and output are listed in the last row of Table 1, if we use information from Figure 13, 17, 21, 28, 45, and 57. we can calculate exactly how many possibilities for certain length of path. Lastly, with the heatmap benchmarks, we can better visualize the biological process in time domain.

To summarize, with the mixed results from characterizations in Table 1 and visualization from heatmaps, we can clearly observe the inherent nature of any models, which the results can be used as the ground truth for future bio-modelers. Also, with fast developing speed of computation algorithms and tools, which proposed in this thesis can be verified and updated as well. So that the biology community can always have the up-to-date benchmarks to measure and compare the performance of different tools.

Table 1 Benchmarked models and their characteristics

	Breast cancer	Angiogenesis	Tcell	ABA	Macrophage	LGL(reduced)	PCC
Elements	20	25	45	54	142	62	261
Edges	51	31	82	76	287	144	373
Total feedback loops	7	0	199	11	136	23283	18112
Positive feedback	7	0	128	4	42	11817	9668
Negative feedback	0	0	71	7	94	11466	8444
Inner loops	4	0	2115	5	344	1462004	293443
Paths between all I/Os	1375	26	3668	529	1231	69991	57450544

5.0 Conclusion and future work

5.1 Conclusion

Researchers in biology communities have developed a huge amount of methods and algorithms to address scientific problems. However, if many approaches exist, biologists usually have a hard time deciding which tool is the best compared to the rest, because a new method always claims its advantages superior to the previous. Accordingly, an ideal solution for this dilemma is to create benchmarks, which stands for golden ground truth, so that researchers can compare different tools in an unbiased fashion.

Biological characteristics are the cornerstone of benchmark metrics to measure the preciseness and completeness of the model. As you can see in Table 1, we carefully select a few specifications, which are the most common metrics that suits the biology community's interest.

5.2 Future work

To better improve the quantity and quality of biological benchmarks, there are several improvements which can be done. First of all, computation time for the large model like PCC is significantly longer than the small model like the breast cancer model. One of the characteristics, such as generating all paths between input and output, takes more than one day. If any hardware implementation such as FPGA acceleration can be designed, this could dramatically lower the benchmarking time. Secondly, except the most common metrics chosen in this paper, which lay

the foundation of biological benchmarking, I believe there should be more interesting and useful metrics can be added and expanded, such as model sensitivity analysis. Last but not least, instead of generating specifications and plots in Jupyter Notebook, an executable graph user interface (GUI) can better present the benchmarked models to all researchers.

Bibliography

[1] K. K. Kummer, T. Kalpachidou, M. Kress, and M. Langeslag, "Signatures of Altered Gene Expression in Dorsal Root Ganglia of a Fabry Disease Mouse Model," (in eng), *Front Mol Neurosci*, vol. 10, p. 449, 2017, doi: 10.3389/fnmol.2017.00449.

[2] R. G. Cooper and E. J. Kleinschmidt, "Benchmarking the firm's critical success factors in new product development," vol. 12, no. 5, pp. 374-391, 1995.

[3] S. Greenberg, E. Mills, B. Tschudi, and P. Rumsey, "Best Practices for Data Centers: Lessons Learned from Benchmarking 22 Data Centers," 2006.

[4] E. Platen, "A benchmark approach to finance," (in English), *Mathematical Finance*, Article|Proceedings Paper vol. 16, no. 1, pp. 131-151, JAN 2006 2006, doi: 10.1111/j.1467-9965.2006.00265.x.

[5] M. Ruhul Amin, "Benchmarking learning outcomes of undergraduate business education," *Benchmarking: An International Journal*, vol. 10, no. 6, pp. 538-558, 2003, doi: 10.1108/14635770310505175.

[6] N. Renaudie, "A Methodology For Parametrizing Discrete Model of Biological Networks," Master degree, Swanson School of Engineering > Electrical and Computer Engineering, University of Pittsburgh, 2018. [Online]. Available: http://d-scholarship.pitt.edu/34202/

[7] D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions," *The Journal of Physical Chemistry*, vol. 81, no. 25, pp. 2340-2361, 1977/12/01 1977, doi: 10.1021/j100540a008.

[8] B. Novak and J. J. Tyson, "A model for restriction point control of the mammalian cell cycle," (in eng), *J Theor Biol*, vol. 230, no. 4, pp. 563-79, Oct 21 2004, doi: 10.1016/j.jtbi.2004.04.039.

[9] J. R. Faeder, M. L. Blinov, and W. S. Hlavacek, "Rule-based modeling of biochemical systems with BioNetGen," (in eng), *Methods Mol Biol*, vol. 500, pp. 113-67, 2009, doi: 10.1007/978-1-59745-525-1_5.

[10] G. Zhou, "Sensitivity Analysis of Discrete Models and Application in Biological Networks," Master degree, Swanson School of Engineering > Electrical Engineering, University of Pittsburgh, 2018. [Online]. Available: http://d-scholarship.pitt.edu/34134/

[11] A. Saadatpour, I. Albert, and R. Albert, "Attractor analysis of asynchronous Boolean models of signal transduction networks," (in eng), *J Theor Biol*, vol. 266, no. 4, pp. 641-56, Oct 21 2010, doi: 10.1016/j.jtbi.2010.07.022.

[12] R. Zhang *et al.*, "Network model of survival signaling in large granular lymphocyte leukemia," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 105, no. 42, pp. 16308-16313, 2008.

[13] Ö. Sahin *et al.*, "Modeling ERBB receptor-regulated G1/S transition to find novel targets for de novo trastuzumab resistance," vol. 3, no. 1, p. 1, 2009.

[14] A. L. Bauer, T. L. Jackson, Y. Jiang, and T. Rohlf, "Receptor cross-talk in angiogenesis: mapping environmental cues to cell phenotype using a stochastic, Boolean signaling network model," (in eng), *J Theor Biol,* vol. 264, no. 3, pp. 838-46, Jun 7 2010, doi: 10.1016/j.jtbi.2010.03.025.

[15] N. Miskov-Zivanov, M. S. Turner, L. P. Kane, P. A. Morel, and J. R. Faeder, "The duration of T cell stimulation is a critical determinant of cell fate and plasticity," (in eng), *Sci Signal,* vol. 6, no. 300, p. ra97, Nov 5 2013, doi: 10.1126/scisignal.2004217.

[16] A. Butchy, "Macrophage model," ed, 2019.

[17] Telmer Cheryl *et al.*, "Computational Modeling of Cell Signaling and Mutations in Pancreatic Cancer," presented at the Artificial Intelligence for Synthetic Biology Conference, Arlington, USA, 2019.

[18] K. Sayed, C. A. Telmer, A. A. Butchy, and N. Miskov-Zivanov, "Recipes for Translating Big Data Machine Reading to Executable Cellular Signaling Models," Cham, 2018: Springer International Publishing, pp. 1-15.

[19] "Standard Performance Evaluation Corporation." https://www.spec.org/spec/ (accessed 2019).

[20] "Embedded Microprocessor Benchmark Consortium." https://www.eembc.org (accessed 2019).

[21] "EEMBC Benchmarks Products." https://www.eembc.org/products/ (accessed.

[22] "EEMBC CoreMark." https://www.eembc.org/coremark/ (accessed.

[23] SPEC® CPU2017 Floating Point Rate Result [Online] Available: https://www.spec.org/cpu2017/results/res2017q4/cpu2017-20171211-01438.html

[24] V. Chelliah *et al.*, "BioModels: ten-year anniversary," *Nucleic Acids Research,* vol. 43, no. D1, pp. D542-D548, 2014, doi: 10.1093/nar/gku1181.

[25] L. A. Chylek, L. A. Harris, C. S. Tung, J. R. Faeder, C. F. Lopez, and W. S. Hlavacek, "Rule-based modeling: a computational approach for studying biomolecular site dynamics

in cell signaling systems," (in eng), *Wiley Interdiscip Rev Syst Biol Med,* vol. 6, no. 1, pp. 13-36, Jan-Feb 2014, doi: 10.1002/wsbm.1245.

[26] K. Gilboy, K. Sayed, N. Sundaram, K. Bocan, and N. Miskov-Zivanov, "A Faster DiSH: Hardware Implementation of a Discrete Cell Signaling Network Simulator," in *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, 27-30 May 2018 2018, pp. 1-5, doi: 10.1109/ISCAS.2018.8350960.

[27] S. A. Kauffman, "Metabolic stability and epigenesis in randomly constructed genetic nets," vol. 22, no. 3, pp. 437-467, 1969.

[28] R. D. A. Rene Thomas, *Biological Feedback*. CRC Press, 1990.

[29] S. L. Harris and A. J. Levine, "The p53 pathway: positive and negative feedback loops," (in eng), *Oncogene,* vol. 24, no. 17, pp. 2899-908, Apr 18 2005, doi: 10.1038/sj.onc.1208615.

[30] S.-Y. Shin *et al.*, "Positive- and negative-feedback regulations coordinate the dynamic behavior of the Ras-Raf-MEK-ERK signal transduction pathway," *Journal of Cell Science,* vol. 122, no. 3, p. 425, 2009, doi: 10.1242/jcs.036319.

[31] S.-Y. Shin, O. Rath, A. Zebisch, S.-M. Choo, W. Kolch, and K.-H. Cho, "Functional roles of multiple feedback loops in extracellular signal-regulated kinase and Wnt signaling pathways that regulate epithelial-mesenchymal transition," (in eng), *Cancer research,* vol. 70, no. 17, pp. 6715-6724, 2010, doi: 10.1158/0008-5472.CAN-10-1377.

[32] J. B.-J. a. G. Gutin, *Digraphs, theory, algorithms and applications*, 2nd ed. Springer Monographs in Mathematics, 2008.

[33] Y.-K. Kwon and K.-H. Cho, "Coherent coupling of feedback loops: a design principle of cell signaling networks," *Bioinformatics,* vol. 24, no. 17, pp. 1926-1932, 2008, doi: 10.1093/bioinformatics/btn337.

[34] J.-R. Kim, Y. Yoon, and K.-H. Cho, "Coupled feedback loops form dynamic motifs of cellular networks," (in eng), *Biophys J,* vol. 94, no. 2, pp. 359-365, 2008, doi: 10.1529/biophysj.107.105106.

[35] D. Gil, J. Law, and T. Murali, "The PathLinker app: Connect the dots in protein interaction networks [version 1; peer review: 1 approved, 2 approved with reservations]," *F1000Research,* vol. 6, no. 58, 2017, doi: 10.12688/f1000research.9909.1.

[36] R. Albert *et al.*, "A new discrete dynamic model of ABA-induced stomatal closure predicts key feedback loops," *PLOS Biology,* vol. 15, no. 9, p. e2003451, 2017.

[37] A. K. Singh, *Digital VLSI Design*. Prentice Hall of India, 2010

[38] K. Sayed, Y. Kuo, A. Kulkarni, and N. Miskov-Zivanov, "DiSH simulator: Capturing dynamics of cellular signaling with heterogeneous knowledge," in *2017 Winter Simulation Conference (WSC)*, 3-6 Dec. 2017 2017, pp. 896-907, doi: 10.1109/WSC.2017.8247841.

[39] S. Li, S. M. Assmann, and R. Albert, "Predicting essential components of signal transduction networks: a dynamic model of guard cell abscisic acid signaling," (in eng), *PLoS Biol*, vol. 4, no. 10, p. e312, Oct 2006, doi: 10.1371/journal.pbio.0040312.

[40] A. Mantovani and A. Sica, "Macrophages, innate immunity and cancer: balance, tolerance, and diversity," (in eng), *Curr Opin Immunol*, vol. 22, no. 2, pp. 231-7, Apr 2010, doi: 10.1016/j.coi.2010.01.009.

[41] I. B. McInnes and G. Schett, "The pathogenesis of rheumatoid arthritis," *New England Journal of Medicine*, vol. 365, no. 23, pp. 2205-2219, 2011.

[42] J. M. Olefsky and C. K. Glass, "Macrophages, inflammation, and insulin resistance," *Annual review of physiology*, vol. 72, pp. 219-246, 2010.

[43] M. C. Bosco, "Macrophage polarization: Reaching across the aisle?," ed: Elsevier, 2019.

[44] D. M. Mosser and J. P. Edwards, "Exploring the full spectrum of macrophage activation," *Nature reviews immunology*, vol. 8, no. 12, p. 958, 2008.

[45] P. J. Murray *et al.*, "Macrophage activation and polarization: nomenclature and experimental guidelines," *Immunity*, vol. 41, no. 1, pp. 14-20, 2014.

[46] M. Fallahi-Sichani, M. El-Kebir, S. Marino, D. E. Kirschner, and J. J. Linderman, "Multiscale computational modeling reveals a critical role for TNF-α receptor 1 dynamics in tuberculosis granuloma formation," *The Journal of Immunology*, vol. 186, no. 6, pp. 3472-3483, 2011.

[47] S. Marino, N. A. Cilfone, J. T. Mattila, J. J. Linderman, J. L. Flynn, and D. E. Kirschner, "Macrophage polarization drives granuloma outcome during Mycobacterium tuberculosis infection," *Infection and immunity*, vol. 83, no. 1, pp. 324-338, 2015.

[48] N. A. Cilfone *et al.*, "Computational modeling predicts IL-10 control of lesion sterilization by balancing early host immunity–mediated antimicrobial responses with caseation during Mycobacterium tuberculosis infection," *The Journal of Immunology*, vol. 194, no. 2, pp. 664-677, 2015.

[49] M. Helmy, J. Gohda, J.-i. Inoue, M. Tomita, M. Tsuchiya, and K. Selvarajoo, "Predicting novel features of toll-like receptor 3 signaling in macrophages," *PLoS One*, vol. 4, no. 3, p. e4661, 2009.

[50] A. Palma, A. S. Jarrah, P. Tieri, G. Cesareni, and F. Castiglione, "Gene regulatory network modeling of macrophage differentiation corroborates the continuum hypothesis of polarization states," *Frontiers in physiology*, vol. 9, p. 1659, 2018.

[51] S. Raza *et al.*, "Construction of a large scale integrated map of macrophage pathogen recognition and effector systems," *BMC systems biology*, vol. 4, no. 1, p. 63, 2010.

[52] S. Raza *et al.*, "A logic-based diagram of signalling pathways central to macrophage activation," *BMC systems biology*, vol. 2, no. 1, p. 36, 2008.

[53] M. Nagasaki *et al.*, "Systems biology model repository for macrophage pathway simulation," *Bioinformatics*, vol. 27, no. 11, pp. 1591-1593, 2011.