Modeling of Yeast Pheromone Pathway using Petri Nets

Abhishek Majumdar
University of Nebraska-Lincoln, majumdar@cse.unl.edu

Follow this and additional works at: http://digitalcommons.unl.edu/computerscidiss
Part of the Computer Engineering Commons

http://digitalcommons.unl.edu/computerscidiss/53

This Article is brought to you for free and open access by the Computer Science and Engineering, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Computer Science and Engineering: Theses, Dissertations, and Student Research by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Modeling of Yeast Pheromone Pathway using Petri Nets

by

Abhishek Majumdar

Presented to the Faculty of
The Graduate College at the University of Nebraska
In Partial Fulfillment of Requirements
For the Degree of Master of Science

Major: Computer Science

Under the Supervision of Professor Stephen D. Scott and
Professor Jitender Singh Deogun

Lincoln, Nebraska

December, 2012
Modeling of Yeast Pheromone Pathway using Petri Nets

Abhishek Majundar, M.S.
University of Nebraska, 2012

Advisors: Stephen D. Scott and Jitender Singh Deogun

Yeast (Saccharomyces cerevisiae) is one of the most widely studied single celled organisms. Mating of yeast cells occur between cells of opposite mating types a and α. Pheromone secretion by a cell alerts the corresponding opposite type cell about its presence and eventually facilitates the process of mating between them. The details of how pheromones affect cells can be studied from the pheromone response pathway in a yeast cell. A response pathway typically depicts the chain of interactions that happens between the different proteins in the cells in response to the pheromone. In this thesis we model the yeast pheromone response pathway using Petri nets and simulate various conditions under which a cell will respond positively to the secreted pheromone. We also take a look at how different proteins in the cell might be able to facilitate the correct functionality of the pathway. The objective of this thesis is to serve as a guideline for performing lab experiments to further explore the yeast pheromone pathway.
ACKNOWLEDGEMENTS

My heart felt gratitude goes to my advisors, Prof. Stephen D. Scott and Prof. Jitender S. Deogun, for their continuous support, guidance and encouragement during the thesis. Without their patience, motivation, guidance, immense knowledge and encouragement it would not have been possible for me to complete the thesis.

I would like to thank Prof. Steven Harris, for his support during the thesis. His expertise has been a guiding star in the development of this thesis.

I would like to thank Dr. Monica Heiner of Brandenburg University of Technology, Cottbus, Germany for her advise on setting up the experiments.
## Contents

1 Introduction ............................................. 6

2 Background ................................................ 9
   2.1 Pheromone Response Process in Yeast ................. 9
   2.2 Petri nets ............................................. 11
      2.2.1 Introduction ....................................... 11
      2.2.2 Definitions ........................................ 12
      2.2.3 Properties ........................................ 14

3 Problem Description ..................................... 18

4 Related Work ............................................. 20

5 Modeling method ......................................... 24
   5.1 Model .................................................. 24
      5.1.1 Modifications ...................................... 25
      5.1.2 Additions .......................................... 27

6 Petri net representation of the pathway ................ 29
   6.1 Model .................................................. 29

7 Experiment and Results .................................. 40
   7.1 Experimental Setup .................................... 40
   7.2 Simulating the Network ................................ 42
      7.2.1 Simulation Experiments ........................... 42
      7.2.2 Results ............................................ 44
      7.2.3 Interpretation of Results ......................... 47
   7.3 Analysis of Experiments ............................... 48
      7.3.1 Development of Decision Trees .................... 48
      7.3.2 Results ............................................ 51
      7.3.3 Interpretation of Results ......................... 55
8 Conclusions
8.1 Discussion of Results ........................................ 58
8.2 Future Work ...................................................... 59

Bibliography .......................................................... 61
Chapter 1

Introduction

Our environment is composed of tightly interlinked complex systems at different levels of magnitude. The smallest biological level of detail is the molecular level of DNA, RNA, proteins and metabolites. All these molecules together form a cell. Cells form tissues. Different tissues constitute the organs of an organism. Different organisms together form the ecosystem. So the relationships between different biological elements at any level of detail can be represented as a network called a biological network [6].

To date, five different types of biological networks have been characterized: transcription factor binding, protein-protein interactions, protein phosphorylation, metabolic interactions and genetic interaction networks. Transcription is the process of copying DNA to mRNA. A transcription factor is a protein which performs this function acting alone or with other agents in a complex either as an activator or a repressor. Protein phosphorylation is a post-translational modification of proteins by addition of phosphate groups. Metabolic interaction networks focus on the mass flow in basic chemical pathways that generate essential components such as amino acids, sugars, lipids etc and the energy required by the biochemical reactions. Gene interaction or protein interaction networks, as the names indicate, reflect the interaction between
the different genes/proteins [16]. Modeling of biological networks helps to come up with an overall view about the state of biological knowledge of the network and also to reveal important properties that might remain hidden without any proper models. Also, these models can be used to predict behaviors about the network which can then be verified with experimental results.

Yeasts are single celled microorganisms in the Fungi kingdom. *Saccharomyces cerevisiae* a particular species of yeast, has been widely studied in genetics and cell biology. It is a simple eukaryote cell which serves as a model for all eukaryotes for the study of fundamental cellular processes like cell cycle, cell division, metabolism, DNA replication, etc. [5] *S. cerevisiae* has both asexual and sexual reproduction. Sexual reproduction takes place between two haploid cells of opposite types a and α. The process of mating is initiated by secretion of pheromone by one of the cells. Receptors on the opposite cell detect the presence of pheromone and initiates a series of protein-protein interactions within the cell which ultimate might facilitate mating. This series of protein-protein interactions in the cell is known as the yeast pheromone pathway. This pathway is a well studied problem where the basic structure and different proteins participating in the pathway are now known. So this is essentially a protein-protein interaction network.

In this thesis we use Petri nets [14] to model this protein interaction network. To this end we propose a model of the pheromone pathway (adopted from Sackmann et al. [15]), modify it according to our needs, and augment it with some additional proteins which we believe might have some accessory role in the process [9]. In absence of consistent real world data of the different kd values of the different protein interactions we generate those randomly from the range \{1, 2, \ldots, 100\} [9]. We simulate the pathway using our model to identify the conditions under which yeast cells will mate. The contributions of the thesis include putting together the model, identifying a set
of rules that determines, whether a cell would successfully mate or not and finally identifying a possible process used by yeast cell to overcome detrimental conditions and successfully mate.
Chapter 2

Background

2.1 Pheromone Response Process in Yeast

In this chapter we describe the process of pheromone binding to its receptor on the cell surface and the subsequent effects of that phenomenon on the cell functionality. The chapter is based on the textbook by H. Madhani [12]. The mating process is initiated, when a yeast cell detects the presence of pheromone secreted by a cell of opposite sex. Mating consists of sequence of processes; activation of gene expression, arrest of cell cycle, polarized growth of shmoo (cellular projection in yeast cell) towards the mating partner, cell-cell fusion and nuclear fusion. There are two cell types in yeast called a and α. They are analogous to egg and sperm cells of animals. These a and α cells can mate to produce an a/α cell. This in turn undergoes meiosis to produce the haploid gametes (child cells) a and α cells. An a cell produces a pheromone known as a-factor whereas an α cell produces a pheromone known as α-factor. An a cell contains the α-factor receptor Ste2 whereas an α cell contains the a-factor receptor Ste3. So a cells can mate with α cells only and vice-versa.

For both Ste2 and Ste3, binding to one of the mating pheromone affects its abil-
ity to interact with an intracellular protein complex known as the heterotrimeric G protein. The G protein is made up of subunits called Gpa1, Ste4 and Ste18. In the remainder of this thesis these subunits will be referred to as Gα, Gβ, and Gγ, respectively. In the G-protein complex, the Gβ and Gγ units form a complex Gβγ, which remain bound to Gα when it is bound to GDP. When pheromone binds to the receptor (Ste2 or Ste3), the receptor interacts with Gα, causing it to replace its GDP with GTP. Gα without its GDP cannot keep the Gβγ complex bound to itself. As a result the Gβγ complex is liberated and goes on to interact with other proteins. Gradually GTP bound to Gα is hydrolyzed to GDP by Gα. It then binds back and inhibits the Gβγ complex.

The Gβγ complex when liberated activates a pathway in which four protein kinases are linked in a series to form a cascade kinase. The Gβγ recruits the protein Ste5 which acts as a scaffold to hold three other proteins Ste11, Ste7 and Fus3. These three proteins activate each other in series by phosphorylating them. So an activated Ste11 phosphorylates Ste7, which becomes active and in turn phosphorylates Fus3. The activated Fus3 then leaves the scaffold and enters the nucleus. This final activation triggers the arrest of the cell cycle as well as the transcription of genes involved in mating. The Ste11 at the top of the kinase is activated by a protein Ste20. The protein Ste20 itself becomes activated when it is in the plasma membrane where it is phosphorylated by a membrane associated monomeric GTPase called Cdc42.

Activated Fus3 plays an important role in both cell cycle arresting as well as the transcription of genes. Activated Fus3 phosphorylates protein Far1 which blocks the cell cycle in G1 phase, to prepare for mating. Fus3 in the nucleus activates the transcription factor Ste12. Ste12 promotes the expression of α-specific and α-specific genes. Normally, in absence of pheromone signal, Ste12 is inhibited by proteins Dig1 and Dig2, which bind to different parts of Ste12 (Dig1 to the DNA-binding domain of
Ste12 and Dig2 to the carboxyl terminus required for DNA-bound Ste12 to activate transcription). Due to pheromone signalling, activated Fus3 phosphorylates Dig1 and Dig2 which in turn release Ste12. The Ste12 is then free to bind and promote the transcription of $\alpha$-specific genes ($\alpha$-sgs) and $\alpha$-specific genes ($\alpha$-sgs).

A prominent feature of mating is the formation of shmoo by the cells towards each other. The pheromone receptors most activated on the cell surface are the ones facing the highest concentration of pheromone. This region contains the highest concentration of $G_{\beta\gamma}$. The $G_{\beta\gamma}$ complex recruits proteins to promote the formation of the shmoo. Far1 binds to the $G_{\beta\gamma}$ complex and recruits three proteins Cdc42, Cdc24 and Bem1 which promote polarized cell growth. Cdc24 activates Cdc42 by inducing transformation of GDP to GTP. The Cdc42 bound to GTP together with Bem1, recruit proteins which promote cell membrane growth such as Bni1 and others. Thus a series of protein-protein interactions beginning with active $G_{\beta\gamma}$ result in the growth of cell membrane towards the highest concentration of pheromone.

To re-enter cell cycle after successful mating or a failed mating attempt, yeast cells have some kind of built-in mechanism which can counter the effects of pheromone exposure on the different proteins. This is typically achieved through various proteins acting in negative feedback loops. Implementation of such loops is beyond the scope of this thesis and hence not discussed here.

### 2.2 Petri nets

#### 2.2.1 Introduction

Petri nets were first proposed by Carl Adam Petri in 1962. Petri nets are a graphical and mathematical tool applicable to many systems. They can be used for describing and modeling systems that can be characterized as concurrent, asynchronous,
distributed, parallel, non-deterministic, and/or stochastic. As a graphical tool they provide the functionality of flowcharts or block diagrams. Petri nets allow the simulation and visualization of dynamic activities of systems by the use of tokens in its modeling. As a mathematical tool they can be used to represent state equations and other mathematical models governing the behaviour of systems. The following discussion is based mostly on the paper by T. Murata [14].

A Petri net is a kind of a directed graph with an initial marking $M_0$. The underlying graph of a Petri net is a directed weighted bipartite graph with two kinds of nodes, places and transitions. Places are represented as circles and transitions are represented as boxes. The arcs are either from places to transitions or vice-versa. The arcs or edges are labeled with weights (positive integers). A unit weight edge is not labeled. A marking $M$ is an assignment of non-negative integers to the places. If a marking assigns integer $x$ to place $p$, then $p$ is said to be marked with $x$ tokens. Pictorially tokens are denoted by dots. Transitions cannot be assigned any integers.

If a Petri net has $m$ places, then a marking is a vector of size $m$. The $p^{th}$ element of $M$, denoted $M(p)$ is the number of tokens in place $p$.

2.2.2 Definitions

In modelling, places represent conditions and transitions represent events. The formal definition of a Petri net is given below.

A Petri net can be defined as a 5-tuple, $PN = (P, T, E, W, M_0)$, where:

$P = \{p_1, p_2, \ldots, p_m\}$ is a finite set of places,

$T = \{t_1, t_2, \ldots, t_n\}$ is a finite set of transitions,

$E \subseteq (P \times T) \cup (T \times P)$ is a set of edges,

$W: E \rightarrow 1, 2, 3, \ldots$ is a weight function,

$M_0: P \rightarrow 0, 1, 2, \ldots$ is the initial marking.
\( P \cap T = \emptyset \) and \( P \cup T \neq \emptyset \).

A Petri net without any initial marking is denoted as \( N \) while that with a given initial marking is denoted by \( (N, M_0) \). Below we define some terminologies related to Petri nets. A preplace of a transition \( t \), denoted \( \text{pre}(t) \) is defined as a place from which there is an incoming edge to the \( t \). Mathematically,

\[
\text{pre}(t) = \{ p \mid (p, t) \in E \}.
\]

Similarly a postplace of a transition \( t \), denoted \( \text{post}(t) \) is defined as a place to which there is an edge from \( t \). Mathematically,

\[
\text{post}(t) = \{ p \mid (t, p) \in E \}.
\]

Similar notations are available for places as well. We define the pre-transition of a place \( p \) as

\[
\text{pre}(p) = \{ t \mid (t, p) \in E \}
\]

and post-transition of a place \( p \) as

\[
\text{post}(p) = \{ t \mid (p, t) \in E \}.
\]

In order to simulate the behaviour of a system, the Petri net is operated using the following firing rules:

1. A transition \( t \) is said to be enabled if each preplace \( p \) of \( t \) has at least \( w(p, t) \) tokens where \( w(p, t) \) is the weight of the edge in between \( p \) and \( t \).

2. An enabled transition may or may not fire (depending on whether the event actually occurs or not)
3. Once transition $t$ has fired, each of its preplace $p$ lose $w(p,t)$ tokens and each of its postplaces $p'$ gains $w(t,p')$ tokens.

![Figure 2.1: A simple Petri net](image1.png)

In Figure 2.1 we show an example of a Petri net with 3 places $p_1$, $p_2$ and $p_3$ and a transition $T_1$. $p_1$ has 10 tokens and $p_2$ has 7 tokens, $w(p_1,T_1) = 5$, $w(p_2,T_1) = 6$ and $w(T_1,p_3) = 9$. By the above mentioned rule 1, $T_1$ is enabled. Thus when $T_1$ fires, the Petri net will change as shown in Figure 2.2. After $T_1$ has fired, $p_1$ has 5 tokens, $p_2$ has 1 token and $p_3$ has 9 tokens (rule 3).

![Figure 2.2: The Petri net of Figure 2.1 after T1 fires](image2.png)

### 2.2.3 Properties

Two types of properties can be studied with a Petri net: *Behavioral properties*, which depend on the initial markings and *Structural properties*, which depend on the topo-
logical structure of the Petri net. We will first discuss some of the behavioral properties and then the structural.

Behavioral Properties

The following properties have been adopted from the paper by T. Murata [14].

1. **Reachability**: A marking $M_n$ in a Petri net is said to be reachable from another marking $M_0$ if there exists a sequence of transition firings that results in transforming the marking $M_0$ to $M_n$. The set of all markings reachable from $M_0$ is denoted by $R(M_0)$.

2. **Boundedness**: A Petri net $(N, M_0)$ is said to be $k$-bounded if the number of tokens in any place does not exceed some positive integer $k$ for any marking reachable from $M_0$.

3. **Liveness**: A Petri net $(N, M_0)$ is said to be live if irrespective of the current marking of the net, any transition can eventually be fired by progressing through some firing sequence.

4. **Reversibility**: A Petri net is said to be reversible if any initial marking $M_0$ can be reached from all markings reachable from $M_0$.

5. **Coverability**: A marking $M$ in a Petri net is said to be coverable if there exists another marking $M' \in R(M_0)$ such that $M'(p) > M_0(p)$ for each place $p$ in the net.

6. **Persistence**: A Petri net $(N, M_0)$ is said to be persistent if, for any two enabled transitions, the firing of one transition will not disable the other.
Structural Properties

Structural properties of a Petri net are the properties which are independent of the initial marking $M_0$. These are dependent on the topology of the net. Before describing the properties, we introduce the concept of incidence matrix in a Petri net. For a Petri net $N$ with $n$ transitions and $m$ places, the incidence matrix $A = [a_{ij}]$ is an $n \times m$ matrix of integers where each matrix entry is given by

$$a_{ij} = a_{ij}^+ - a_{ij}^-,$$

where $a_{ij}^+ = w(i,j)$ is the weight of the edge from transition $i$ to place $j$ and $a_{ij}^- = w(j,i)$ is the weight of the edge from place $j$ to transition $i$. Some papers use $A^T$ as the incidence matrix instead of $A$. The structural properties of a Petri net can be represented in terms of its incidence matrix $A$ and its associated homogeneous equations or inequalities. For the following properties, the net is assumed to be free of self loops. For any vector $z$, the $i^{th}$ element of $z$ is denoted by $z(i)$.

1. **Structural Liveness**: A Petri net $N$ is said to be **structurally live** if there exists a live initial marking for $N$.

2. **Controllability**: A Petri net $N$ is said to be **completely controllable** if any marking is reachable from any other marking.

3. **Structural Boundedness**: A Petri net is said to be **structurally bounded** if it is bounded for any finite initial marking $M_0$.

4. **Conservativeness**: A Petri net is said to be (partially) **conservative** if there exists a positive integer $y(p)$ for (some) every place $p$ such that the weighted sum of the tokens $M^T y = M_0^T y$ is a constant for every $M \in R(M_0)$ and for any fixed initial marking $M_0$. 
5. **Repetitiveness**: A Petri net is said to (partially) repetitive if there exists an initial marking $M_0$ and a firing sequence $\varpi$ such that (some) every transition occurs often in $\varpi$.

6. **Consistency**: A Petri net is said to be (partially) consistent if there exists a marking $M_0$ and a firing sequence $\varpi$ from $M_0$ to $M_0$ such that (some) every transition occurs at least once in $\varpi$.

**Invariants**

- P-invariant: An $m$-vector $y$ of integers is called a *P-invariant* if $Ay = 0$.

- T-invariant: An $n$-vector $x$ of integers is called a *T-invariant* if $A^Tx = 0$.

The set of places corresponding to a non-zero entry in a P-invariant is called the support of an invariant. A support is said to be minimal if no other support is contained in it. An invariant $y$ is said to be a minimal invariant if there exists no other invariant $y_1$ such that $y_1(p) \leq y(p)$ for all $p$. For a given minimal support of an invariant there exists a unique invariant corresponding to the minimal support. This is known as minimal support invariant. Linear combinations of minimal-support invariants yield new invariants.
Chapter 3

Problem Description

The yeast pheromone response pathway is one of the most well-studied biological pathways. Over the years, ample experiments have been performed which have allowed biologists to piece together the different components of the pathway. As a result we now have a working knowledge of how the pathway functions, the different proteins that take part in this pathway and their respective roles. However, several questions still remain unanswered. One particular question is: how does the cell dynamically adapt the pathway to continue mating under severe environmental changes or under mutation (which might result in the loss of functionality of some proteins known to participate in the pheromone pathway). That is, it is not entirely known how the cell compensates, uses other proteins to contribute towards functioning of the pathway if the original pathway is compromised.

This thesis attempts to answer the above mentioned question. To achieve this, we first propose a model to simulate the pheromone pathway. Then we look at the following two questions.

1. Given the model of the pheromone response pathway, under what conditions does the cell respond positively, i.e., mate?
2. What kind of perturbations in the cell would result in changing a negative response to a positive one?

In our model, the “conditions” mentioned in question 1 typically refers to the different edges weights between the different components of the pathway. Different combinations of the values of the edge weights represent different environmental conditions faced by the cell. The best way to answer question 1 would be to come up with a set of rules that can be used to predict whether a cell will be able to successfully mate or not. “Perturbations” mentioned in question 2 refers to possible methods employed by the cell so that it can mate. We hypothesise that one such method might be to make use of accessory proteins who otherwise are not so prominent in the pheormone pathway. Using appropriate amounts of proteins other than the core pathway component proteins can be a possible compensation method used by the cell to facilitate mating. To answer question 2 we try to come up with a list of accessor proteins and figure out their functionality. The combined answers of questions 1 and 2 are expected to provide a clearer understanding of this process.
Chapter 4

Related Work

In this chapter we discuss related work. We survey some of the papers in which a Petri net approach has been used to model biological networks.

Sachmann et al. [15] provide a systemic modeling method of signal transduction pathways in terms of Petri net components. The authors present a process of representing the following three different cases of a signal transduction model.

Case 1: A substance $A$ does not lose its activity by interacting with a second substance $B$.

Case 2: A substance $C$ triggers several reactions that are independent of each other.

Case 3: A substance changes state from being phosphorylated to being unphosphorylated and vice versa.

Case 1 indicates phosphorylation reactions between different proteins in a network. Case 2 describes participation of a protein in multiple independent reactions. Both cases are implemented by using read arcs (bidirectional edges between places and transitions) in their Petri net representations. Case 3 indicates the different states of a protein, which is implemented in form of a sub-network. Having described these, the authors propose the following simple steps for representing a signal pathway. First,
translate the biological components into logical structures like conjunction, disjunction, exclusive disjunction and implication. Second, translate the logical structures in corresponding Petri net forms. Finally, assimilate the Petri net components to form a whole network. Our work uses the modeling approach used by this paper and forms the basic structure of our model on the model provided in this paper.

Chaouiya [7] provides an overview of the different types of Petri net models available and their uses in modeling different types of biological networks. A Coloured Petri Net (CPN) uses different colored tokens to represent different data types. They also allow assigning expressions to the edges. For firing of a transition, the corresponding expression on the incoming edges to the transition are to be satisfied. Similarly, to determine the output of a transition, the expression on its outgoing edge needs to be calculated. CPNs have been used to simulate enzymatic reaction chains.

A Stochastic Petri Net (SPN) allows time delay functions to be associated with the firing of transitions. This allows SPNs to incorporate uncertainty and noise into its modeling. SPNs have been used to model and simulate stochastic molecular interactions. Hybrid Petri Nets (HPNs) allow modeling of both discrete and continuous processes. For discrete events they have discrete places with tokens and transitions while for continuous processes they have continuous places with real variables and continuous transitions which can fire continuously at a given rate. To increase their expressiveness, they also have read arcs and inhibitor arcs (a transition will fire if the concentration(s) of its pre-place(s) are zero). Hybrid Function Petri Nets (HFPNs) have some added functionality in addition to those of HPNs. Continuous transition firing rates can be made dependent on the value of its pre-places and the weights of the arcs can be defined as a function of the marking of the connected places. HPNs have been used for quantitative modeling and simulation of gene regulatory networks. Several regulated metabolic pathways have been modelled using HFPNs.
The paper concludes with discussion that the use of Petri nets for modeling metabolic reactions is relatively intuitive. CPNs are capable of representing gene interaction or signal transduction networks, while HPNs support representation of a wide range of molecular mechanism.

Hardy and Robillard [8] also discuss the different types of Petri nets extensions used for analysis, modeling and simulation of molecular biology networks. They identify two categories of goals of Petri net biological modeling: qualitative and quantitative analysis. Qualitative analysis is the analysis of the different biological properties while quantitative analysis is the simulation of system dynamics. For qualitative analysis, a suitable Petri net extension should be chosen that allows its various different properties to be determined. For quantitative analysis, a Petri net representation with sufficient modeling power should be chosen. The authors provide a summary of different modeling goals for the different Petri net extensions. CPNs are good for analysis of biological system properties while SPNs and HFPNs are good for simulation purposes. For quantitative analysis of a biological system, kinetic parameters like reaction rates and stoichiometric quantities of reactants are necessary. In this thesis since no such data was available, we used the basic Petri net structure for our quantitative analysis. In the future, pending availability of data, we plan to upgrade our model to a HFPN or something similar.

Monica et al. [10] demonstrate a generalized approach towards modeling and analysis of biological pathways using Petri nets. For model validation they utilize one of the basic behavioral properties of Petri nets: T-Invariants. To demonstrate the effectiveness of their approach they model the apoptosis (cell death) process. For modeling, they follow a few basic rules. Biochemical substances are represented by places and any relation between biochemical substances are represented by transitions with corresponding arcs. Signal transduction events and enzyme catalytic reactions are
modelled using test arcs. For model validation, the authors use T-Invariants, which in the context of biological networks are sets of reactions reproducing a given system state. To facilitate validation, the paper changes the model into an empty Petri net where all input output nodes are transitions. For such a net, the T-Invariants are the sets of transitions reproducing the empty marking. Using T-Invariant analysis they are successfully able to account for all the known basic behaviors of the process hence validating their model.
Chapter 5

Modeling method

5.1 Model

We use Petri nets to model the pheromone response pathway. We adopt and modify the approach mentioned in the papers by Heiner et al. [10] and Sachmann [15]. We represent each protein as a place in the Petri net and each interaction as a transition. Hence a reaction $R$ between proteins $p_A$ and $p_B$ to produce protein $p_C$ can be represented as the following net structure. In Figure 5.1, the places A and B represent the

![Petri net representation of a reaction process](image)

Figure 5.1: Petrinet representation of a reaction process
proteins $p_A$ and $p_B$ and transition R represents the reaction between A and B that forms protein $p_C$ (place C). The tokens in each place represent the initial concentration of each protein. Also, the edge weights between R and the pre-places represent concentration thresholds to be exceeded by individual protein. The firing rules for transitions in Figure 5.1 are given by:

1. Number of tokens from A $\geq W_A$ \textbf{AND}

2. Number of tokens from B $\geq W_B$

Using this representation as the basic unit of construction, all reactions in the pathway can be modeled and the full pathway can be obtained by combining these individual reaction representations. Such a model is already available in the paper by Sachmann et al [15]. We base our model on this available network structure [15]. We make several changes to this model to suit our solution approach. In the following section we describe those changes and modifications and the rationale behind them.

\textbf{5.1.1 Modifications}

In a Petri net, a transition can fire if it is \textit{enabled}. By that definition in Figure 5.1, for reaction R to fire and form compound C, both A and B must have at least $W_A$ and $W_B$ tokens each, respectively. So essentially this is a conjunction rule where all pre-places must be present and have some pre-requisite number of tokens for a reaction to occur. We propose a method to convert all such conjunction rules into corresponding disjunction rules.

We know that the reaction between two or more proteins takes place if the corresponding $kd$ value exceeds a certain threshold value. As an example, say proteins A and B interact to form C only when the strength of their interaction ($kd$ value)
is more than, say, $W_{AB}$. The conventional Petri net representation as shown in Figure 5.1 does not allow for this $kd$ value concept to be implemented properly. For this reason we have come up with the disjunctive representation of reactions in Petri nets. The same reaction involving $A$, $B$ and $C$ can be represented in Figure 5.2.

To Figure 5.1, we have added two dummy post-transitions $D_A$ and $D_B$ to $A$ and $B$ respectively and another preplace $AB$ to transition $R$. The weight of the edge from place $AB$ to transition $R$ is $W_{AB}$ which makes sure that $R$ will fire only if the number of tokens coming from its pre-place exceeds the threshold value. The firing rules of the different transitions in Figure 5.2 are as follows:

1. $D_A$: Number of tokens from $A \geq W_A$

2. $D_B$: Number of tokens from $B \geq W_B$

3. $R$: Number of tokens from $AB \geq W_{AB} \Rightarrow W_A + W_B \geq W_{AB}$

Comparing Figure 5.1 and Figure 5.2 we see that they essentially represent the same reaction. This notion can be generalized to a reaction with $n$ participants. Assume a reaction $R$ in which $n$ proteins $p_1, p_2, \ldots, p_i, \ldots, p_n$ interact to produce a compound $q$. Let $W_R$ be the threshold value that needs to be exceeded by the cumulative weight/concentration of the reactants for $R$ to proceed. Let $w_1, w_2, \ldots, w_i, \ldots, w_n$
be the weights/concentrations of the reacting proteins respectively. The disjunctive representation of this reaction can be drawn as follows. For each protein $p_i$, construct a dummy transition $D_i$ where both the incoming and outgoing edges have weight $w_i$. Construct a pre-place $S$ to $R$ such all dummy transitions become pre-transitions to $S$. Now let the edge weight between $S$ and $R$ be $W_R$. Thus the firing rules for the different transitions become:

1. $D_i$: Number of tokens from $p_i \geq w_i \ \forall \ i \in \{1.....n\}$

2. $R$: Number of tokens from $S \geq W_R \Rightarrow \sum_{i=1}^{n} w_i \geq W_R$

The firing rule for $R$ imposes the threshold condition that we want to be included while representing reactions. Also it gives one added advantage. It allows the assumption of weighted sum of the components for successful execution of the reaction. That is, all reactants no longer need to be present in its entirety for the reaction to proceed. As long as the cumulative concentration of the reactants exceeds the reaction threshold, it will work, which makes more sense as in nature that is what happens [9].

5.1.2 Additions

We augment the above model by adding more proteins that exist in the yeast cell, and are believed (not yet proved) to play some role in the pheromone response process. The proteins taking part in the pheromone response process are known to interact physically with other proteins. We looked up the *Saccharomyces* Genome Database [3] for all proteins which are known to have physical interaction with the core protein components (Ste5, Ste11, Ste7, Ste20, Ste50, Fus3, Dig1, Dig2, Ste12, Sst2, Far1, Cdc24, Cdc42, Bem1, Ste2, Ste3, Ste4, Ste18, GPA1 and Tec1) of the pathway which are already known to us. For each of those proteins we then look up their properties to see if they are known to play some role in pheromone interaction pathways. We
retain all proteins who are known to have some role related to pheromone exchange or those whose functions are unknown. The rationale of adding these proteins is to give the model enough options so that it can to some extent simulate how the cell might use these proteins in the pheromone pathway.
Chapter 6

Petri net representation of the pathway

6.1 Model

We use the pheromone response pathway structure provided by Sachmann et al [15]. In that representation each protein component is represented by a place and each reaction by a transition. We modify this model so as to incorporate the idea of requisite kd values for the different reactions. To this end we transform the preplaces of all transitions to a single place (marked red in our model) which has inputs from different reactant places. To conform to the Petri Net design, we add a dummy transition to each reactant place as shown in the previous section. Only for transitions with Ste-type proteins as pre-places are left unchanged. The benefit of having a single pre-place to a transition which originally required several pre-places is that it emphasizes the notion of weighted cumulative concentration of the reactants. Table 6.2 [15] gives a list of all the protein components and their symbols used in our model. Table 6.4 [15] gives a list of all the transitions, their symbols and biological reactions
that they represent. We then add more proteins that are known to interact with various component proteins of the pheromone pathway. We obtain these additional proteins from the yeast genome database. The steps followed are described below. First, for each of the 20 protein components, in the core pathway namely Ste5, Ste11, Ste7, Ste20, Ste50, Fus3, Dig1, Dig2, Ste12, Sst2, Far1, Cdc24, Cdc42, Bem1, Ste2, Ste3, Ste4, Ste18, GPA1 and Tec1, we list all proteins that are known to interact with them physically. From this list we select only those proteins that are known to participate in the pheromone pathway reactions. Some proteins from this list overlap with the list of core component proteins. So they are not used in the model. So finally the list thus obtained (Table 6.5) contains 37 new proteins, which we will add to the pathway. We take these 37 additional proteins and add them to our network structure in the following manner. For each protein \(i\) which has \(j\) as a neighboring protein, we make \(i\) participate in all the reactions in which \(j\) is a reactant. In terms of our model, \(i\) becomes a preplace to all the post-transitions of \(j\). Table 6.6 lists transitions and their preplaces, in other words, the reactants responsible for each reaction.

After adding the additional proteins we add regulatory edges (colored blue) in Figure 6.1 in the network to control the order in which transitions may fire in the network. We define regulatory edges as bidirectional edges of weight one between a place and a transition which makes sure that the transition cannot fire until that place has at least one token. Also, bidirectionality ensures that the token content of the place is not affected by the firing of the transition. We illustrate this with the help of Figure 6.1. In Figure 6.1, reaction \(T_1\) produces compound \(P_1\), which participates in reaction \(T_2\). Protein \(P_0\) participates in reaction \(T_2\) which in turn produces \(P_2\). In the figure the bidirectional edge (blue edge) between \(P_1\) and \(T_2\) is a regulatory edge that makes sure that \(T_2\) will not fire until \(P_1\) is produced by \(T_1\) irrespective of the amount of \(P_0\) present.
Figure 6.1: Example of Regulatory edge

Figure 6.2 illustrates the full structure of our representation of the pheromone pathway. In the structure, all places starting with \( p \) refer to the core protein components of the pathway listed in Table 6.2 and all places starting with \( a \) refer to the newly added proteins listed in Table 6.5. The places marked red represent the weighted sum listed in Table 6.7. All transitions starting with \( t \) are the transitions listed in Table 6.4. All logical nodes (connector which allow a node to participate in multiple activities in different areas) are marked in grey. The edges marked in blue are the regulatory edges introduced in the structure.
Figure 6.2: Full Petri net representation of the pathway
Table 6.1: Places of the model

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Place Name</th>
<th>Biological species</th>
</tr>
</thead>
<tbody>
<tr>
<td>p1</td>
<td>alpha-factor</td>
<td>pheromone released by an MAT&lt;sup&gt;a&lt;/sup&gt; cell in the surroundings</td>
</tr>
<tr>
<td>p2</td>
<td>Ste2 receptor</td>
<td>mating pheromone receptor of the modelled MAT&lt;sup&gt;a&lt;/sup&gt; cell</td>
</tr>
<tr>
<td>p3</td>
<td>receptor_factor_complex</td>
<td>complex consisting of the α factor and the Ste2 receptor</td>
</tr>
<tr>
<td>p4</td>
<td>receptor_complex</td>
<td>the above named complex is activated by a conformation change</td>
</tr>
<tr>
<td>p5</td>
<td>trimer_bound_to_receptor</td>
<td>heterotrimeric G protein, which is coupled to the Ste2 receptor</td>
</tr>
<tr>
<td>p6</td>
<td>G&lt;sub&gt;alpha&lt;/sub&gt; GTP</td>
<td>dissociated Gα subunit (exchange of GDP to GTP in this monomer)</td>
</tr>
<tr>
<td>p7</td>
<td>G&lt;sub&gt;beta&lt;/sub&gt; gamma_dimer</td>
<td>G-protein Gβγ subunits in a dimer form</td>
</tr>
<tr>
<td>p8</td>
<td>Cdc24</td>
<td>Cdc24, i.e., guanine nucleotide exchange factor of Cdc42</td>
</tr>
<tr>
<td>p9</td>
<td>Cdc42(at pm)</td>
<td>Cdc42 located at the plasma membrane</td>
</tr>
<tr>
<td>p10</td>
<td>Ste20</td>
<td>protein kinase Ste20</td>
</tr>
<tr>
<td>p11</td>
<td>Ste5(scaffold)</td>
<td>Ste5, acting as a scaffold protein</td>
</tr>
<tr>
<td>p12</td>
<td>Ste5/Ste11</td>
<td>protein complex consisting of ste7 and Fus3</td>
</tr>
<tr>
<td>p13</td>
<td>Fus3</td>
<td>MAP kinase Fus3</td>
</tr>
<tr>
<td>p14</td>
<td>Ste7/Fus3</td>
<td>protein complex consisting of Ste7 and Fus3</td>
</tr>
<tr>
<td>p15</td>
<td>MAPK_complex</td>
<td>MAPK complex consisting of Ste5,Ste11,Ste7 and Fus3</td>
</tr>
<tr>
<td>p16</td>
<td>Ste20_at_pm</td>
<td>Ste20 located at the plasma membrane, i.e., near the MAPK complex</td>
</tr>
<tr>
<td>p17</td>
<td>complex2</td>
<td>as complex1, but Ste11 is activated additionally</td>
</tr>
<tr>
<td>p18</td>
<td>complex3</td>
<td>as complex2, but Ste7 is activated additionally</td>
</tr>
<tr>
<td>p19</td>
<td>complex4</td>
<td>as complex3, but Fus3 is activated additionally</td>
</tr>
<tr>
<td>p20</td>
<td>Fus3PP</td>
<td>dissociated Fus3 in the activated form</td>
</tr>
<tr>
<td>p21</td>
<td>complex_without_Fus3</td>
<td>as complex4, but without Fus3</td>
</tr>
<tr>
<td>p22</td>
<td>repr_complex</td>
<td>complex containing Ste12 repressed by Fus3 or Kss1</td>
</tr>
<tr>
<td>p23</td>
<td>Dig1/Dig2</td>
<td>Ste12 inhibitors, i.e., cofactors for the repression</td>
</tr>
<tr>
<td>p24</td>
<td>free_Ste12</td>
<td>Ste12 released out of the repression complex</td>
</tr>
<tr>
<td>p25</td>
<td>Ste12</td>
<td>activated transcription factor Ste12</td>
</tr>
<tr>
<td>p26</td>
<td>Msg5</td>
<td>phosphatase Msg5 being able to deactivate Fus3 or Kss1</td>
</tr>
<tr>
<td>p27</td>
<td>Fus3.dephos</td>
<td>deactivated Fus3</td>
</tr>
<tr>
<td>p28</td>
<td>other genes</td>
<td>pheromone regulated genes encoding mating related cell responses</td>
</tr>
<tr>
<td>p29</td>
<td>Bar1_in_nucleus</td>
<td>synthesised protease Bar1 located in the nucleus</td>
</tr>
<tr>
<td>p30</td>
<td>Bar1</td>
<td>Bar1 secreted in the cell environment</td>
</tr>
<tr>
<td>p31</td>
<td>inactive_Far1</td>
<td>synthesised Far1 located in the nucleus in an inactive form</td>
</tr>
<tr>
<td>p32</td>
<td>Far1</td>
<td>Far1 activated by phosphorylation</td>
</tr>
<tr>
<td>p33</td>
<td>Far1_in_cytosol</td>
<td>active Far1 located in the cytosol</td>
</tr>
<tr>
<td>Page</td>
<td>Description</td>
<td>Details</td>
</tr>
<tr>
<td>------</td>
<td>------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>p34</td>
<td>Sst2_in_nucleus</td>
<td>synthesised Sst2 located in the nucleus in an inactive form</td>
</tr>
<tr>
<td>p35</td>
<td>phos_Sst2</td>
<td>Sst2 activated by phosphorylation</td>
</tr>
<tr>
<td>p36</td>
<td>Sst2</td>
<td>active Sst2 located in the cytosol</td>
</tr>
<tr>
<td>p37</td>
<td>inactive_component</td>
<td>complex labelled for degradation by phosphorylation</td>
</tr>
<tr>
<td>p38</td>
<td>phos_Kss1</td>
<td>MAP kinase Kss1 activated by phosphorylation</td>
</tr>
<tr>
<td>p39</td>
<td>unphos_Kss1</td>
<td>inactive Kss1</td>
</tr>
<tr>
<td>p40</td>
<td>Akr1</td>
<td>protein Akr1 located at plasma membrane</td>
</tr>
<tr>
<td>p41</td>
<td>Yck1/Yck2_at_pm</td>
<td>kinases Yck1/Yck2 being able to label the Ste2 for degradation</td>
</tr>
<tr>
<td>p42</td>
<td>inactive_receptor</td>
<td>receptor labelled for ubiquitination and endocytosis</td>
</tr>
<tr>
<td>p43</td>
<td>Ste11</td>
<td>protein kinase Ste11</td>
</tr>
<tr>
<td>p44</td>
<td>Ste50</td>
<td>protein kinase Ste50</td>
</tr>
<tr>
<td>p45</td>
<td>Bem1</td>
<td>protein Bem1</td>
</tr>
<tr>
<td>p46</td>
<td>Ste7</td>
<td>protein kinase Ste7</td>
</tr>
<tr>
<td>Symbol</td>
<td>Transition Name</td>
<td>Biological Event</td>
</tr>
<tr>
<td>---------</td>
<td>-----------------------------------------------------</td>
<td>----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>t1</td>
<td>MATalpha_cell(surroundings)</td>
<td>MATα cell secretes its mating pheromone</td>
</tr>
<tr>
<td>t2</td>
<td>binding_factor_to_receptor</td>
<td>α-factor binds to the Ste2 receptor</td>
</tr>
<tr>
<td>t3</td>
<td>receptor_synthesis</td>
<td>synthesis of the cell surface Sst2</td>
</tr>
<tr>
<td>t4</td>
<td>receptor_conformation_change</td>
<td>conformation change of the receptor</td>
</tr>
<tr>
<td>t5</td>
<td>division(in_alpha_subunit:GDP)→GTP</td>
<td>dissociation of the Ga subunit</td>
</tr>
<tr>
<td></td>
<td></td>
<td>of the G-protein</td>
</tr>
<tr>
<td>t6</td>
<td>hydrolysis_GTP→GDP</td>
<td>hydrolysis reassociates Go with Gβγ</td>
</tr>
<tr>
<td>t7</td>
<td>interact_through_Far1</td>
<td>Gβγ interacts Far1 transmitted with Cdc24</td>
</tr>
<tr>
<td>t8</td>
<td>Cdc42:GDP→GTP</td>
<td>Cdc24 activates Cdc42</td>
</tr>
<tr>
<td>t9</td>
<td>active_Cdc42_constitutive_at_pm</td>
<td>constitute active Cdc42 attending the processes</td>
</tr>
<tr>
<td>t10</td>
<td>Ste20_input</td>
<td>source of Ste20</td>
</tr>
<tr>
<td>t11</td>
<td>Ste20Activated</td>
<td>Cdc42 at plasma membrane and Bem1 activates Ste20</td>
</tr>
<tr>
<td>t12</td>
<td>Ste5_input</td>
<td>source of Ste5</td>
</tr>
<tr>
<td>t13</td>
<td>Ste5 binds_Ste11</td>
<td>Ste5 binds Ste11</td>
</tr>
<tr>
<td>t14</td>
<td>Fus3_synth</td>
<td>synthesis of kinase Fus3</td>
</tr>
<tr>
<td>t15</td>
<td>Fus3 binds_Ste7</td>
<td>Ste7 binds Fus3</td>
</tr>
<tr>
<td>t16</td>
<td>complex-formation</td>
<td>Ste5/Ste11 binds Ste7/Fus3</td>
</tr>
<tr>
<td>t17</td>
<td>Ste20 Phos_Ste11</td>
<td>phosphorylation of Ste11 by Ste20</td>
</tr>
<tr>
<td>t18</td>
<td>Ste11 Phos_Ste7</td>
<td>phosphorylation of Ste7 by Ste11</td>
</tr>
<tr>
<td>t19</td>
<td>Ste7 Phos_Fus3</td>
<td>phosphorylation of Fus3 by Ste7</td>
</tr>
<tr>
<td>t20</td>
<td>Fus3PP-release</td>
<td>release of activation Fus3 out of the MAPK complex</td>
</tr>
<tr>
<td>t21</td>
<td>binding_free_Fus3</td>
<td>remaining MAPK complex binds Fus3</td>
</tr>
<tr>
<td>t22</td>
<td>Ste12_inhibit_phos</td>
<td>phosphorylation of Ste12 inhibitors Dig1/Dig2 by Fus3PP</td>
</tr>
<tr>
<td>t23</td>
<td>Ste12-release</td>
<td>release of Ste12 out of the repression complex</td>
</tr>
<tr>
<td>t24</td>
<td>Ste12_phos</td>
<td>phosphorylation of Ste12 by Fus3PP</td>
</tr>
<tr>
<td>t25</td>
<td>transcr_activation</td>
<td>transcription activation of pheromone regulated genes</td>
</tr>
<tr>
<td>t26</td>
<td>Fus3PP_dephos</td>
<td>dephosphorylation of Fus3PP by Msg5</td>
</tr>
<tr>
<td>t27</td>
<td>repression_through_Fus3</td>
<td>Ste12 repression through inactive Fus3 and Dig1/Dig2</td>
</tr>
<tr>
<td>t28</td>
<td>cell_fusion</td>
<td>processes leading to the fusion of the two haploid cells</td>
</tr>
<tr>
<td>t29</td>
<td>transport_out_of_cell</td>
<td>Bar1 transport into the cell environment</td>
</tr>
<tr>
<td>t30</td>
<td>factor_destruction</td>
<td>Bar1 transmitted destruction of the α-factor</td>
</tr>
<tr>
<td>t31</td>
<td>Far1_phos</td>
<td>phosphorylation of Far1 by Fus3PP</td>
</tr>
<tr>
<td>t32</td>
<td>cell_cycle_arrest in_G1</td>
<td>Far1 caused arrest in the cell cycle phase G1</td>
</tr>
<tr>
<td>t33</td>
<td>transport_out_of_cell</td>
<td>Bar1 transport into the cell environment</td>
</tr>
<tr>
<td>-----</td>
<td>-------------------------</td>
<td>----------------------------------------</td>
</tr>
<tr>
<td>t34</td>
<td>Sst2_phos</td>
<td>phosphorylation of Far1 by Fus3PP</td>
</tr>
<tr>
<td>t35</td>
<td>transport_out_of_nucleus</td>
<td>Sst2 transport out of the nucleus</td>
</tr>
<tr>
<td>t36</td>
<td>accelerated_hydr_GTP\rightarrow_GDP</td>
<td>accelerated hydrolysis reassociates the G-protein</td>
</tr>
<tr>
<td>t37</td>
<td>Ste11_neg_phos</td>
<td>Fus3PP labels the MAPK complex at Ste11 for degradation</td>
</tr>
<tr>
<td>t38</td>
<td>degradation</td>
<td>degradation of the MAPK complex</td>
</tr>
<tr>
<td>t39</td>
<td>Ste7_neg_phos</td>
<td>Fus3PP labels the MAPK complex at Ste7 for degradation</td>
</tr>
<tr>
<td>t40</td>
<td>Ste7_phos_Kss1</td>
<td>phosphorylation of Kss1 by Ste7</td>
</tr>
<tr>
<td>t41</td>
<td>accelerated-dephos-Kss1</td>
<td>deactivation of phosphorylation Kss1 by Fus3PP</td>
</tr>
<tr>
<td>t42</td>
<td>Kss1_dephos</td>
<td>dephosphorylation of phosphorylated Kss1 by Msg5</td>
</tr>
<tr>
<td>t43</td>
<td>repression_through_Kss1</td>
<td>Ste12 repression through inactive Kss1 by Msg5</td>
</tr>
<tr>
<td>t44</td>
<td>tech_input</td>
<td>techinal:the repressed Ste12 complex assumed to be present</td>
</tr>
<tr>
<td>t45</td>
<td>Akr_synthesis</td>
<td>synthesis of Akr1</td>
</tr>
<tr>
<td>t46</td>
<td>Akr1_binds_Yck1_Yck2</td>
<td>Akr1 binds Yck1_Yck2</td>
</tr>
<tr>
<td>t47</td>
<td>receptor_phos</td>
<td>labelling of Ste2 for degradation</td>
</tr>
<tr>
<td>t48</td>
<td>ubiquit_endocytosis</td>
<td>ubiquitination and endocytosis of the receptor</td>
</tr>
<tr>
<td>Symbol</td>
<td>Protein Name</td>
<td>Neighboring Components</td>
</tr>
<tr>
<td>--------</td>
<td>--------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>a1</td>
<td>CBK1</td>
<td>STE5, STE20, STE50</td>
</tr>
<tr>
<td>a2</td>
<td>PTC1</td>
<td>STE5, STE20</td>
</tr>
<tr>
<td>a3</td>
<td>CLA4</td>
<td>STE11, CDC24, CDC42, BEM1</td>
</tr>
<tr>
<td>a4</td>
<td>DSE1</td>
<td>STE11, STE4</td>
</tr>
<tr>
<td>a5</td>
<td>HOG1</td>
<td>STE11, STE7, STE50</td>
</tr>
<tr>
<td>a6</td>
<td>PBS2</td>
<td>STE11, BEM1</td>
</tr>
<tr>
<td>a7</td>
<td>SHO1</td>
<td>STE11, STE20, STE50, SST2, CDC24</td>
</tr>
<tr>
<td>a8</td>
<td>SPA2</td>
<td>STE11, STE7</td>
</tr>
<tr>
<td>a9</td>
<td>SPH1</td>
<td>STE11, STE7</td>
</tr>
<tr>
<td>a10</td>
<td>RGA2</td>
<td>STE20, CDC24, CDC42, BEM1</td>
</tr>
<tr>
<td>a11</td>
<td>CLN2</td>
<td>STE20, DIG1, DIG2, FAR1</td>
</tr>
<tr>
<td>a12</td>
<td>ENT2</td>
<td>CDC24, STE20</td>
</tr>
<tr>
<td>a13</td>
<td>EXO84</td>
<td>STE20, BEM1</td>
</tr>
<tr>
<td>a14</td>
<td>BOI1</td>
<td>STE20, FUS3, DIG1, DIG2, CDC24, CDC42, BEM1</td>
</tr>
<tr>
<td>a15</td>
<td>CDC28</td>
<td>STE20, FAR1, BEM1</td>
</tr>
<tr>
<td>a16</td>
<td>GIC1</td>
<td>STE50, CDC42</td>
</tr>
<tr>
<td>a17</td>
<td>GIC2</td>
<td>STE50, CDC24, CDC42</td>
</tr>
<tr>
<td>a18</td>
<td>BN1</td>
<td>FUS3, CDC42</td>
</tr>
<tr>
<td>a19</td>
<td>MPT5</td>
<td>FUS3, SST2</td>
</tr>
<tr>
<td>a20</td>
<td>KDX1</td>
<td>TEC1, DIG1, DIG2, STE12</td>
</tr>
<tr>
<td>a21</td>
<td>KSS1</td>
<td>STE5, STE11, STE7</td>
</tr>
<tr>
<td>a22</td>
<td>WHI3</td>
<td>TEC1, SST2, STE2</td>
</tr>
<tr>
<td>a23</td>
<td>BZZ1</td>
<td>DIG1, DIG2</td>
</tr>
<tr>
<td>a24</td>
<td>HMLALPHA1</td>
<td>DIG1, DIG2, STE12</td>
</tr>
<tr>
<td>a25</td>
<td>HYM1</td>
<td>DIG1, DIG2</td>
</tr>
<tr>
<td>a26</td>
<td>YCK2</td>
<td>DIG2, STE3</td>
</tr>
<tr>
<td>a27</td>
<td>RSR1</td>
<td>CDC42, BEM1, CDC24</td>
</tr>
<tr>
<td>a28</td>
<td>SEC15</td>
<td>CDC24, BEM1</td>
</tr>
<tr>
<td>a29</td>
<td>EXO70</td>
<td>CDC42, BEM1</td>
</tr>
<tr>
<td>a30</td>
<td>SEC3</td>
<td>CDC42, BEM1</td>
</tr>
<tr>
<td>a31</td>
<td>RHO1</td>
<td>BEM1, STE4</td>
</tr>
<tr>
<td>a32</td>
<td>SEC6</td>
<td>BEM1, STE2</td>
</tr>
<tr>
<td>a33</td>
<td>AKR1</td>
<td>BEM1, STE2</td>
</tr>
<tr>
<td>a34</td>
<td>DIB1</td>
<td>STE7, DIG1</td>
</tr>
<tr>
<td>a35</td>
<td>YHR131C</td>
<td>STE20, FUS3</td>
</tr>
<tr>
<td>a36</td>
<td>BDF2</td>
<td>STE20, FUS3</td>
</tr>
<tr>
<td>a37</td>
<td>SAS10</td>
<td>STE20, FUS3</td>
</tr>
<tr>
<td>a38</td>
<td>RBS1</td>
<td>DIG1, DIG2</td>
</tr>
<tr>
<td>a39</td>
<td>YJR003C</td>
<td>DIG1, DIG2</td>
</tr>
<tr>
<td>a40</td>
<td>AXL2</td>
<td>CDC24, CDC42, BEM1</td>
</tr>
<tr>
<td>a41</td>
<td>BEM4</td>
<td>STE20, CDC24, CDC42</td>
</tr>
<tr>
<td>Transitions</td>
<td>Meaning</td>
<td>Pre-places</td>
</tr>
<tr>
<td>-------------</td>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>t2</td>
<td>α-factor binds to the Ste2 receptor</td>
<td>s2 and p2</td>
</tr>
<tr>
<td>t4</td>
<td>conformation change of the receptor</td>
<td>p3</td>
</tr>
<tr>
<td>t5</td>
<td>dissociation of the Gα subunit of the G-protein</td>
<td>s5</td>
</tr>
<tr>
<td>t6</td>
<td>hydrolysis reassociates Gα with Gβγ</td>
<td>s6</td>
</tr>
<tr>
<td>t7</td>
<td>Gβγ interacts Far1 transmitted with Cdc24</td>
<td>s7</td>
</tr>
<tr>
<td>t8</td>
<td>hydrolysis reassociates Gα with Gβγ</td>
<td>p8 and s8</td>
</tr>
<tr>
<td>t11</td>
<td>Cdc42 at plasma membrane, Ste20 and Bem1 activates Ste20</td>
<td>p9, p10 and s11</td>
</tr>
<tr>
<td>t13</td>
<td>Ste5 binds Ste11</td>
<td>p11, p43 and s13</td>
</tr>
<tr>
<td>t15</td>
<td>Ste7 binds Fus3</td>
<td>p46 and s15</td>
</tr>
<tr>
<td>t16</td>
<td>Ste5/Ste11 binds Ste7/Fus3</td>
<td>p12 and p14</td>
</tr>
<tr>
<td>t17</td>
<td>phosphorylation of Ste11 by Ste20</td>
<td>p15, p16, p44 and s17</td>
</tr>
<tr>
<td>t18</td>
<td>phosphorylation of Ste7 by Ste11</td>
<td>p17</td>
</tr>
<tr>
<td>t19</td>
<td>phosphorylation of Fus3 by Ste7</td>
<td>p18</td>
</tr>
<tr>
<td>t20</td>
<td>release of activation Fus3 out of the MAPK complex</td>
<td>p19</td>
</tr>
<tr>
<td>t21</td>
<td>remaining MAPK complex binds Fus3</td>
<td>p21 and s21</td>
</tr>
<tr>
<td>t22</td>
<td>phosphorylation of Ste12 inhibitors Dig1/Dig2 by Fus3PP</td>
<td>s22</td>
</tr>
<tr>
<td>t23</td>
<td>release of Ste12 out of the repression complex</td>
<td>p23 and s23</td>
</tr>
<tr>
<td>t24</td>
<td>phosphorylation of Ste12 by Fus3PP</td>
<td>p24 and s24</td>
</tr>
<tr>
<td>t25</td>
<td>transcription activation of pheromone regulated genes</td>
<td>p25 and s25</td>
</tr>
<tr>
<td>t26</td>
<td>dephosphorylation of Fus3PP by Msg5</td>
<td>s26</td>
</tr>
<tr>
<td>t27</td>
<td>Ste12 repression through inactive Fus3 and Dig1/Dig2</td>
<td>p25 and s27</td>
</tr>
<tr>
<td>t28</td>
<td>processes leading to the fusion of the two haploid cells</td>
<td>p28</td>
</tr>
<tr>
<td>t29</td>
<td>Bar1 transport into the cell environment</td>
<td>p29</td>
</tr>
<tr>
<td>t30</td>
<td>Bar1 transmitted destruction of the α-factor</td>
<td>p1 and p30</td>
</tr>
<tr>
<td>t31</td>
<td>phosphorylation of Far1 by Fus3PP</td>
<td>s31</td>
</tr>
<tr>
<td>t32</td>
<td>Far1 caused arrest in the cell cycle phase G1</td>
<td>s32</td>
</tr>
<tr>
<td>t33</td>
<td>Bar1 transport into the cell environment</td>
<td>s32</td>
</tr>
<tr>
<td>t34</td>
<td>phosphorylation of Far1 by Fus3PP</td>
<td>s34</td>
</tr>
<tr>
<td>t35</td>
<td>Sst2 transport out of the nucleus</td>
<td>s35</td>
</tr>
<tr>
<td>t36</td>
<td>accelerated hydrolysis reassociates the G-protein</td>
<td>s36</td>
</tr>
<tr>
<td>t37</td>
<td>Fus3PP labels the MAPK complex at Ste11 for degradation</td>
<td>p17 and s37</td>
</tr>
<tr>
<td>t38</td>
<td>degradation of the MAPK complex</td>
<td>p37</td>
</tr>
<tr>
<td>t39</td>
<td>Fus3PP labels the MAPK complex at Ste11 for degradation</td>
<td>p18 and s39</td>
</tr>
<tr>
<td>t40</td>
<td>phosphorylation of Kss1 by Ste7</td>
<td>p18 and s40</td>
</tr>
<tr>
<td>t41</td>
<td>deactivation of phosphorylation Kss1 by Fus3PP</td>
<td>s41</td>
</tr>
<tr>
<td>t42</td>
<td>dephosphorylation of phosphorylated Kss1 by Msg5</td>
<td>s42</td>
</tr>
<tr>
<td>t43</td>
<td>Ste12 repression through inactive Kss1 by Msg5</td>
<td>p25 and s43</td>
</tr>
<tr>
<td>t46</td>
<td>Akr1 binds Yck1/Yck2</td>
<td>p40</td>
</tr>
<tr>
<td>t47</td>
<td>labelling of Ste2 for degradation</td>
<td>s47</td>
</tr>
<tr>
<td>t48</td>
<td>ubiquitination and endocytosis of the receptor</td>
<td>p42</td>
</tr>
<tr>
<td>Name</td>
<td>Constituents</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>--------------</td>
<td></td>
</tr>
<tr>
<td>s2</td>
<td>p1,a22,a26,a32,a33</td>
<td></td>
</tr>
<tr>
<td>s5</td>
<td>p4,p5</td>
<td></td>
</tr>
<tr>
<td>s6</td>
<td>p6,p7</td>
<td></td>
</tr>
<tr>
<td>s7</td>
<td>p7,p33,a11,a15</td>
<td></td>
</tr>
<tr>
<td>s8</td>
<td>p45,a3,a7,a10,a12,a28,a27,a17,a14,a40,a41,a6,a15,a13,a31,a30,a32</td>
<td></td>
</tr>
</tbody>
</table>
| s11  | p7,p45,a1,a2,a3,a6,a7,a5,a10,a11,a12,a13,a14,a15,  
a35,a36,a37,a41,a16,a17,a18,a27,a29,a30,a31,a32,a40 |
| s13  | p7,a1,a2,a3,a6,a7,a4,a5,a8,a9,a21 |
| s15  | p13,a21,a34,a5,a8,a9 |
| s17  | a1,a2,a3,a6,a7,a5,a10,a11,a12,a13,a14,a15,a35,a36,a37,a41 |
| s21  | p13         |
| s22  | p20,p22,a14,a10,a19,a35,a36,a37 |
| s23  | a14,a25,a26,a38,a39,a11,a20,a21,a23,a24 |
| s24  | p20,a14,a10,a19,a35,a36,a37 |
| s26  | p26,p20,a14,a10,a19,a35,a36,a37 |
| s27  | p27,s20,s24 |
| s31  | p31,p20,a14,a18,a19,a21,a22,a7 |
| s32  | p32,a11,a15 |
| s34  | p20,p34     |
| s36  | p7,a19,a21,a22,a7 |
| s37  | p20,a14,a18,a19,a35,a36,a37 |
| s39  | p20,a14,a18,a19,a35,a36,a37 |
| s40  | p39         |
| s41  | p20,p38,a14,a18,a19,a35,a36,a37 |
| s42  | p26,p38     |
| s47  | p41,p4      |
Chapter 7

Experiment and Results

7.1 Experimental Setup

The Petri net model as described in the previous chapter can be visualized using the Snoopy 2.0 tool [2]. Using the export feature offered by Snoopy, we converted the graphical representation of the model into ANDL (Abstract Net Definition Language) format [11]. It is basically a textual representation of the network. The places in the network are represented as variables. The variables are initialized with numbers indicating the initial token content of the corresponding places. For instance, \( V_i = 5 \) in the Petri net means that there is a place named \( V_i \) with 5 tokens in it. Each transition is described as result of the changes in its preplaces and postplaces. For instance,

\[
T_1: [p_1 + 34] \& [p_3 + 23] \& [p_1 - 10]
\]

means transition \( T_1 \) has place \( p_1 \) as its \textit{preplace} and \( p_2 \) and \( p_3 \) as its \textit{postplaces}. For \( T_1 \) to fire, \( p_1 \) must have at least 10 tokens. When \( T_1 \) fires, 10 tokens are removed from \( p_1 \) and 34 and 23 tokens are placed into \( p_2 \) and \( p_3 \), respectively.

Using the ANDL description, we developed a Java program that generates ran-
dom networks for the model. In the program the places are represented by an array of variables. Each transition is represented by an object of a Rule class. Once all the transitions have been created, they are stored in a list called rules. As explained in Chapter 5, due to the absence of real world data about the kd values for the different reactions in the pathway, we generate all the edge weights in our model randomly. The range of values for the edge weights used in our experiments is between 1 and 100 (extremities included). The places representing the components α-factor, Ste2-receptor, Ste20, Ste5, Fus3, Akr1, Ste11, Ste7, Ste50 and Bem1 were provided with initial concentration values. Let ψ represent the set of these 10 core component proteins. All places representing the additional components were also provided with initial concentration values. Let λ represent the set of all 41 additional protein components in our model. From here on, the above mentioned places will be referred to as initial-places. For a given value of concentration of all the proteins in sets ψ and λ, the network is simulated. For each transition in the list rules, it is checked whether the transition is enabled or not. If yes, it is fired. After a single pass through the entire list, it is checked whether the transition producing Ste12 has fired or not. If yes, then the pathway has responded successfully and the resultant concentration values of the different proteins are recorded. If not, the entire list is traversed repeatedly. A counter is kept to avoid an infinite loop. After each pass through the list rules, the counter is incremented. However, if after a pass through the entire list, no transition fires, then the network has reached dead state. So the counter is not incremented and the execution is terminated. The execution is also terminated if the counter exceeds a pre-fixed value.
7.2 Simulating the Network

7.2.1 Simulation Experiments

To simulate the pathway, we carry out three different experiments, each of which is discussed below. For the yeast pheromone pathway, apart from the structure of the pathway, exact kd values for each reaction are not known. From the literature, it can be seen that some experiments do provide possible kd values for some reactions. However, such values cannot be used in a generic way because they are specific to particular experiments. We have assumed that the value of kd for each reaction lies within the range \{1, 2, \ldots, 100\} [9]. In absence of real life data, we generate the kd value for each reaction randomly within the range \{1, 2, \ldots, 100\} i.e. we assign weights to the different edges in the network structure randomly from \{1, 2, \ldots, 100\}. The values allowed for each edge are discrete as Petri nets do not allow interchange of fractional tokens. For each experiment, the range of values of concentration allowed for the proteins in set \(\psi\) is \{1, 2, \ldots, 100\} (since Petri nets only allow integer number of tokens to be exchanged). The range of values for proteins in set \(\lambda\) vary in each experiment. Also in the simulation, values of all elements in each set \(\psi\) or \(\lambda\) change together. That is, when one protein in set \(\psi\) has a concentration value of 10, all the other proteins in \(\psi\) are also given the same value. The same is done for \(\lambda\). In the rest of the thesis when we say “value for \(\psi\)” we mean the value of the initial concentration of the proteins in \(\psi\); similarly “value for \(\lambda\)” means the value of the initial concentration of the proteins in \(\lambda\). In biological context, when we are simulating a network with its randomly generated edge weights, the edge weights represent different conditions the cell is subjected to while it tries to respond to pheromone.

1. **Experiment 1**: The range of values of initial concentration for the proteins in \(\lambda\) is set to be between 1 and 10. For any network, for a given value of initial
concentration of proteins in $\psi$ and for a given value of initial concentration of proteins in $\lambda$, it is checked if the transition that produces protein Ste12 has fired or not. Production of Ste12 indicates that mating will happen and we call such an output a positive response. On the other hand, if Ste12 is not produced, it indicates that the cell will not mate and we call this output a negative response by the network. Based on the range of values allowed for both $\psi$ and $\lambda$ there are in total 1000 combinations (100 values from the set $\psi$ times 10 values from the set $\lambda$) of initial concentration values of proteins in $\psi$ and in $\lambda$. We generate 13962 networks and check for the response of the pathway in each of them. The objective of Experiment 1 is to identify conditions (i.e., different edge weights) under which the cell responds positively to the pheromone pathway.

2. **Experiment 2**: We take the 13962 networks generated in Experiment 1, and isolate the networks based on their responses. The ones which gave negative response are put in set $\neg$, while the ones with positive response are put in set $\pos$. We again run the simulation on each of the networks in $\neg$ but now we set the range of concentration of the proteins in $\lambda$ to be $\{11, 12, \ldots, 20\}$. For each network we check, which combination of values for elements in $\psi$ and $\lambda$ yield positive response. If the response is positive, the amount of Ste12 produced, and the concentrations of the core and additional proteins are recorded. The objective of Experiment 2 is to test if the cell can overcome the conditions which made it respond negatively in Experiment 1, by using more concentration of proteins in the set $\lambda$.

3. **Experiment 3**: To have a better understanding of which proteins in $\lambda$ play a more significant role in the pathway, we divide the set $\lambda$ into sets $\sigma$ and $\varsigma$ such that $\lambda = \varsigma \cup \sigma$ and $\sigma \cap \varsigma = \emptyset$. The proteins CBK1, PTC1, DSE1, SPA2, SPH1,
MPT5, KDX1, HYM1, DIB1, YHR131c, BDF2, SAS10, RBS1 and YJR003c from $\lambda$ are placed in $\sigma$. The rest are placed in $\varsigma$. We hypothesise that the proteins in $s$ contribute more to the pheromone pathway than the ones in $\varsigma$ and hence consider them to be more significant in their role in the pathway [9]. To simulate this we set the range of values for the concentration of those proteins to be $\{11, 12, \ldots, 20\}$. For the proteins in $\varsigma$, the range is set to be $\{1, 2, \ldots, 10\}$. For all networks in set $\textit{pos}$ from Experiment 2, we run the simulation and look for positive responses. So the objective of Experiment 3 is to test the above mentioned hypothesis.

7.2.2 Results

1. **Result of Experiment 1**: From the 13962 generated networks, 11816 networks gave negative response. That is, for all 1000 combinations of values of initial concentrations of the proteins in $\psi$ and $\lambda$, in each of the above mentioned 11816 networks, the transition which results in the production of protein Ste12 did not fire. The remaining 2146 networks gave positive a response. The output of networks giving positive response are of 2 types.

(a) A network starts giving positive response when the value for $\psi$ is $\geq$ some value $x \in \{1, 2, \ldots, 100\}$ and the value for $\lambda \geq 1$. For instance, if a network starts giving positive response when the value for $\psi$ is 74 and the value for $\lambda$ is 1, it means that, for this particular network with its set of edge weights (hence worth called \textit{configuration} of the network), as soon the as value for $\psi$ exceeds 74, it will give a positive response irrespective of the concentrations of the proteins in $\lambda$.

(b) A network starts giving positive response when the value for $\psi$ is $\geq$ some
$x \in \{1, 2, \ldots, 100\}$ and the value for $\lambda$ exceeds some value $y \in \{2, \ldots, 10\}$.

For instance, if a network starts giving a positive response when the value for $\psi$ is 74 and the value for $\lambda$ is 5, that means, for this particular network with its corresponding configuration to respond positively, it is not sufficient that the values for $\psi$ become 74; The value for $\lambda$ also needs to exceed value 5.

2. **Result of Experiment 2**: Out of the 11816 networks, 10840 networks still gave negative response. The remaining 976 networks responded positively. That is, out of these 976 networks, each one started giving positive responses when the value for $\psi$ is $\geq$ some value $x \in \{1, 2, \ldots, 100\}$ and the value for $\lambda$ exceeds some value $y \in \{12, \ldots, 20\}$. That is, by increasing the initial concentration level of the proteins in $\lambda$ these networks changed their response from negative in Experiment 1 to positive in this experiment. So this means for these 976 networks, the additional proteins in $\lambda$ play a significant role in deciding how the network responds to the pathway. Changing a prior negative response to a positive one indicates that these proteins might potentially be able to compensate for the lack of some of the core protein components in the pathway if present in sufficient amount.

3. **Result of Experiment 3**: Based on the output of each network, the networks can be classified into 3 categories.

(a) **CS**: This class CS (Class Same) represents those networks that gave positive responses in both Experiments 2 and 3 using the same combination of values for its proteins. That is, if a network gave a positive response in Experiment 2 with values $x$ as the value for $\psi$ and $y$ as the value for $\lambda$, it gives positive response in Experiment 3 as well with the same combination
of values; \( x \) as the value for \( \psi \) and \( y \) as the value for \( \sigma \). For instance, if a network in CS gave a positive response in Experiment 2 when the value for \( \psi \) exceeded 74 and the value for \( \lambda \) exceeded 5, it gives a positive response in Experiment 3 when the value for \( \psi \) exceeded 74 and the value for \( \sigma \) exceeded 5. Out of the 976 networks (from \( pos \)) used for this experiment 352 of them were placed in class \( CS \) because of their output.

(b) CD: This class CD (Class Different) represents those networks which gave positive responses in both Experiments 2 and 3 but using the different combination of values for its proteins. For instance if in Experiment 2, the network had initial concentration values \( x \) for the proteins in \( \psi \) and \( y \) for those in \( \lambda \), in Experiment 3 it has \( x \) as initial concentration value for proteins in \( \psi \) and \( z \) for those in \( \sigma \) where \( y \neq z \). Such a network is placed in class \( CD \). Out of the 976 networks, 76 of them were placed in class \( CD \).

(c) CN: This class CN (Class Negative) represents those networks which gave positive responses in Experiment 2 but now give negative response in Experiment 3. 548 networks from set \( pos \) gave negative response and were placed in class \( CN \).

The distribution of the 3 classes in the set \( pos \) is shown in Figure 7.1.
7.2.3 Interpretation of Results

1. Experiment 1: Networks which give a positive response when the value for $\psi$ is $\geq$ some value $x \in \{1, 2, \ldots, 100\}$ and the value for $\lambda \geq 1$ indicate that for these networks with their corresponding set of edge weights, the additional proteins in $\lambda$ play no significant role in controlling their responses. The response is based solely on the initial concentration of the core component proteins in $\psi$. Networks which start giving a positive response when the value for $\psi$ is $\geq$ some value $x \in \{1, 2, \ldots, 100\}$ and the value for $\lambda \geq$ some $y$ where $y \in \{2, \ldots, 10\}$ indicate that for these networks with their given configuration, depend on the additional proteins in $\lambda$ for modulating their response to the pheromone pathway. That is, for these networks it is the additional proteins in $\lambda$ which makes the response positive when the value for $\psi$ is not sufficient. In a biological context, such networks show that under those conditions the yeast cell uses the proteins in $\lambda$ to facilitate mating. Networks with negative responses indicate the conditions under which a cell will not mate for any combination of initial concentrations of its different proteins.

2. Experiment 2: The 976 networks which start responding positively indicate that the amount of concentration for proteins in $\psi$ or $\lambda$ allowed in Experiment 1 was not sufficient for them to give a positive response. So the cell compensated by using more amounts of those additional proteins in $\lambda$ to facilitate mating. The increase of the range of allowable values for $\lambda$ by us, simulate the cell using more concentration of proteins than what it was using in Experiment 1. These networks support our hypothesis that the cell probably uses one or more additional proteins to respond favorably to the pheromone pathway when it is unable to produce a positive response using just the core component proteins.
3. Experiment 3: Networks in class CS tell us that for these networks with their corresponding configurations the set of proteins in $\sigma$ play a more significant role in the pheromone pathway than the rest of the proteins in $\varsigma$. This indicates that a particular network does not require higher concentrations of all the proteins in $\lambda$ to change its response from negative to positive. The proteins in $\sigma$ are alone capable of doing so. So these networks represent conditions under which the cell rely more on the proteins in $\sigma$ than those in $\varsigma$ to facilitate a change in response from negative to positive.

### 7.3 Analysis of Experiments

#### 7.3.1 Development of Decision Trees

In order to identify reasons that might determine whether a network responds positively or negatively we use decision trees to identify important attributes in the network. “Decision tree learning is a method for approximating discrete-valued target functions, in which the learned function is represented by a decision tree. Learned trees can also be re-represented as sets of if-then rules to improve human readability” [13]. “Decision trees classify instances by sorting them down the tree from the root to some leaf node, which provides the classification of the instance. Each node in the tree specifies a test of some attribute of the instance, and each branched descending from that node corresponds to one of the possible values for this attribute” [13]. We use Weka 3.6 (Waikato Environment for Knowledge Analysis) [4] software for this purpose. We consider each edge in the network as its different attributes. Figure 7.2 shows an example of a small portion of the decision tree generated by our data. Each node or attribute is actually an edge weight from a place to a transition in the network or vice versa. In the tree the values of the weights for edges from places to
transitions are written as negative integers. They should be read as positive values (this is dependent of how attributes were provided to the decision tree learner). So the node a31TOea31 is the value of the edge weight between place a31 and transition ea31 in the network and while checking its value should be considered compared with 21 instead of -21. The very first node at the top is the root of the tree. Each line represents a level in the tree. The straight lines continuing down a node ends at a node with the same attribute but with different checking condition. We start traversing the tree from the root, i.e., the very first node in the tree. If the value of that attribute satisfies that condition we go down that branch i.e. to the next level (next line in the figure). Else we follow the straight line from that node, to its alternate condition. The straight line below node a31TOea31 goes down to a node, where it is checked if attribute a31TOea31 is greater than 21. So at every node, the corresponding condition is checked and a branch is selected accordingly. After traversing the tree from the root, at the fourth level the value of attribute t2TOp3 is checked. If the value is $\leq 34$, then a leaf is reached. That is the decision tree
concludes by looking at the different attribute values in the trees from root to node t2T0p3, that the network will respond negatively. The tree also mentions that how many networks it has correctly and incorrectly classified at that leaf. According to this tree, it has correctly classified 1825 networks as having negative responses and 9 networks incorrectly. at level 4. In this manner the entire tree can be read to predict the response of a network.

The performance of any learning method depends on well it classifies a dataset that it has not encountered before. Cross validation is a method which allows to improve the performance of the learner (in this case the decision tree). The idea is to divide up the entire dataset randomly into a training set (dataset used for training the learner) and a testing set (dataset used for testing the learner). The performance of the learner is then evaluated on this testing set. 10-fold cross validation is a variation of cross validation where the data set is divided into 10 sets. Each time one of the 10 sets is used as a testing set and the remaining 9 become the training set. The whole process is repeated 10 times [1].

1. **Experiment 4**: We take the output of Experiment 1 in Section 7.2.1 and divide the output into two classes $P$ and $N$. Networks that give positive responses are put in class $P$ while the ones with negative response are put in class $N$. For each network, each of its edge weights is listed as an attribute for that network followed by its class $P$ or $N$. Once the list is completed for all 13962 networks, it is given to the J48 decision tree program implemented by Weka 3.6 [4] as an input. A 10-fold cross validation is carried out to get a better estimate of the performance of the decision tree. We compare the different nodes at each level of a decision tree across all the ten trees generated by Weka. This is done to look for attributes which get tested most often (in more than 5 out of 10 trees) at the same level. We look at the first 4 levels starting from the root of each
tree.

2. **Experiment 5**: We take the output of Experiment 2 in Section 7.2.1 and divide the output into two classes \( P \) and \( N \) based on their response as mentioned in Experiment 4. We create a dataset by listing each edge weight of each network followed by their corresponding classes. The dataset is fed to the J48 tree in Weka and 10-fold cross validation is carried out. We compare the nodes at each level across all the 10 trees for the first 4 levels for look for common attributes that get tested often (in more than 5 out of 10 trees) at the same level across all trees.

3. **Experiment 6**: We divide the output of Experiment 3 in Section 7.2.1 into 3 classes \( CS \), \( CD \) and \( CN \), based on their individual responses. These 3 classes are the same ones that we described in Experiment 3 of Section 7.2.1. Once all the networks have been classified, a data set describing the attribute and class of each network is created as mentioned above. The data set is fed to J48 of Weka and a 10-fold cross validation is carried out. We compare the nodes at each level across all the 10 trees for the first 4 levels for look for common attributes that get tested often (in more than 5 out of 10 trees) at the same level across all trees.

### 7.3.2 Results

Figure 7.4 gives the summary of the classification result for **Experiment 4**. From the summary we can see that the tree correctly classifies a network approximately 78% of the time. The confusion matrix tells us that the decision tree has more efficiency (10298/11816 \( \sim \) 87.15%) in correctly predicting a network with negative response than a network with positive response (662/2146 \( \sim \) 30.84%). We identify the most
commonly compared nodes across all 10 trees generated for the first 4 levels. They are listed in Figure 7.5 and Figure 7.6. Each color represents the most common node compared in that level across all 10 trees. Figure 7.5 contains nodes common in the first 3 levels while Figure 7.6 shows nodes common in the fourth level of the trees. From these figures the nodes can be arranged in an increasing order of importance in the following manner. Higher level indicates more importance, more number also adds to the significance of a node. Fig 7.3 gives the color coding scheme followed in the tables. Table 7.1 shows this. Table 7.1 tells us which nodes are examined first by the decision tree while classifying a given network.

Table 7.1: Hierarchy of nodes in Experiment 4

<table>
<thead>
<tr>
<th>Node</th>
<th>Biological Meaning</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>a4Toea4</td>
<td>Minimum concentration of DSE1 required for it to participate in the pathway</td>
<td>7</td>
</tr>
<tr>
<td>t6TOp5</td>
<td>Amount of trimer bound to receptor due to hydrolysis of GTP</td>
<td>10</td>
</tr>
<tr>
<td>p3TOt4</td>
<td>Minimum concentration of receptor-factor complex required for conformation change of the receptor</td>
<td></td>
</tr>
<tr>
<td>a31TOea31</td>
<td>Minimum concentration of RHO1 required for it to participate in the pathway</td>
<td>11</td>
</tr>
<tr>
<td>ea4TOs6</td>
<td>Edge weight between dummy transition for DSE1 and weighted compound s6</td>
<td></td>
</tr>
<tr>
<td>t2TOp3</td>
<td>Amount of receptor-factor complex formed due to binding of $\alpha$ factor to receptor</td>
<td>13</td>
</tr>
<tr>
<td>t2TOp3</td>
<td>Amount of receptor-factor complex formed due to binding of $\alpha$ factor to receptor</td>
<td>50</td>
</tr>
<tr>
<td>e31TOs6</td>
<td>Edge weight between dummy transition for RHO1 and weighted compound s6</td>
<td>51</td>
</tr>
</tbody>
</table>
Figure 7.4: Summary of Experiment 4

Figure 7.5: Common nodes observed in Experiment 4

Figure 7.7 gives the summary of the classification result for Experiment 5. From the summary we can see that the tree correctly classifies a network approximately 87% of the time. The confusion matrix tells us that the decision tree has an efficiency of 209/976 ∼ 21.41% in correctly predicting a positively responding network and an efficiency of 10072/10840 ∼ 92.91% in predicting a network with negative response. We identify the most commonly compared nodes across all 10 trees generated for the first 4 levels. They are listed in Figure 7.8 and Figure 7.9. Each color represents the most common node compared in that level across all 10 trees. Figure 7.8 contains nodes common in the first 3 levels while Figure 7.9 shows nodes common in the
Figure 7.6: Common nodes observed in Experiment 4

fourth level of the trees. Like before we arrange the nodes in an increasing order of importance following the previously mentioned convention. Table 7.2 shows this. It tells us which nodes are examined first by the decision tree while classifying a given network. Figure 7.10 gives the summary of the classification result for Experiment 6. From the summary we can see that the tree correctly classifies a network approximately 94% of the time. The confusion matrix tells us about the efficiency of the decision tree in correctly classifying instances of CS, CN and CD. We identify the most commonly compared nodes across all 10 trees generated for the first 4 levels. For this particular experiment and dataset common nodes were found.
across the trees for the first three levels only. They are listed in Figure 7.11. Each color represents the most common node compared in that level across all 10 trees. We arrange the nodes in an increasing order of importance following the previously mentioned convention. Table 7.3 shows this. It tells us which nodes are examined first by the decision tree while classifying a given network.

### 7.3.3 Interpretation of Results

The tables of most commonly tested attributes obtained from Experiment 4, 5 and 6 reveal that those are the nodes or attributes that drive the decision into classifying a network. So in the context of biological network, these attributes probably represent important conditions that regulate the cell response to pheromone.
Table 7.2: Hierarchy of nodes in Experiment 5

<table>
<thead>
<tr>
<th>Node</th>
<th>Biological Meaning</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>a31TOea31</td>
<td>Minimum concentration of RHO1 required for it to participate in the pathway</td>
<td>21</td>
</tr>
<tr>
<td>a4TOea4</td>
<td>Minimum concentration of DSE1 required to participate in the pathway</td>
<td>21</td>
</tr>
<tr>
<td>s6TOt6</td>
<td>kd value for the reaction: Hydrolysis of G-protein</td>
<td>22</td>
</tr>
<tr>
<td>p3TOt4</td>
<td>Minimum concentration of receptor-factor complex required for conformation change of the receptor</td>
<td>28</td>
</tr>
<tr>
<td>s6TOt6</td>
<td>kd value for the reaction: Hydrolysis of G-protein</td>
<td>46</td>
</tr>
<tr>
<td>t2TOp3</td>
<td>Amount of receptor-factor complex formed due to binding of $\alpha$ factor to receptor</td>
<td>34</td>
</tr>
<tr>
<td>t2TOp3</td>
<td>Amount of receptor-factor complex formed due to binding of $\alpha$ factor to receptor</td>
<td>27</td>
</tr>
<tr>
<td>ea4TOs6</td>
<td>Edge weight between dummy transition for DSE1 and weighted compound s6</td>
<td></td>
</tr>
<tr>
<td>t17TOp17</td>
<td>Concentration of MAPK complex after Ste20 phosphorylates Ste11</td>
<td></td>
</tr>
<tr>
<td>ea31TOs6</td>
<td>Edge weight between dummy transition for RHO1 and weighted compound s6</td>
<td>23</td>
</tr>
</tbody>
</table>

Table 7.3: Hierarchy of nodes in Experiment 6

<table>
<thead>
<tr>
<th>Node</th>
<th>Biological Meaning</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>a4TOea4</td>
<td>Minimum concentration of DSE1 required to participate in the pathway</td>
<td>21</td>
</tr>
<tr>
<td>s6TOt6</td>
<td>kd value for the reaction: Hydrolysis of G-protein</td>
<td></td>
</tr>
<tr>
<td>ea4TOs6</td>
<td>Edge weight between dummy transition for DSE1 and weighted compound s6</td>
<td></td>
</tr>
</tbody>
</table>
Figure 7.10: Summary of Experiment 6

<table>
<thead>
<tr>
<th>Correctly Classified Instances</th>
<th>10083</th>
<th>94.3394 %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incorrectly Classified Instances</td>
<td>602</td>
<td>8.6696 %</td>
</tr>
<tr>
<td>Kappa statistic</td>
<td>0.1017</td>
<td></td>
</tr>
<tr>
<td>Mean absolute error</td>
<td>0.0481</td>
<td></td>
</tr>
<tr>
<td>Root mean squared error</td>
<td>0.1809</td>
<td></td>
</tr>
<tr>
<td>Relative absolute error</td>
<td>91.7861 %</td>
<td></td>
</tr>
<tr>
<td>Root relative squared error</td>
<td>124.8394 %</td>
<td></td>
</tr>
<tr>
<td>Total Number of Instances</td>
<td>10685</td>
<td></td>
</tr>
</tbody>
</table>

### Detailed Accuracy By Class

<table>
<thead>
<tr>
<th>TP Rate</th>
<th>FP Rate</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
<th>ROC Area</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.114</td>
<td>0.021</td>
<td>0.143</td>
<td>0.114</td>
<td>0.127</td>
<td>0.404</td>
<td>CS</td>
</tr>
<tr>
<td>0.975</td>
<td>0.87</td>
<td>0.968</td>
<td>0.975</td>
<td>0.971</td>
<td>0.404</td>
<td>CN</td>
</tr>
<tr>
<td>0.014</td>
<td>0.005</td>
<td>0.019</td>
<td>0.014</td>
<td>0.017</td>
<td>0.21</td>
<td>CD</td>
</tr>
</tbody>
</table>

Weighted Avg. | 0.943 | 0.839 | 0.937 | 0.943 | 0.94 | 0.402 |

### Confusion Matrix

<table>
<thead>
<tr>
<th>a</th>
<th>b</th>
<th>c</th>
<th>--- classified as ---</th>
</tr>
</thead>
<tbody>
<tr>
<td>36</td>
<td>275</td>
<td>5</td>
<td>a = CS</td>
</tr>
<tr>
<td>208</td>
<td>9996</td>
<td>46</td>
<td>b = CN</td>
</tr>
<tr>
<td>8</td>
<td>60</td>
<td>1</td>
<td>a = CD</td>
</tr>
</tbody>
</table>

Figure 7.11: Common nodes observed in Experiment 6
Chapter 8

Conclusions

8.1 Discussion of Results

The simulation experiments revealed 3 kinds of result. Experiment 1 tells us about the different conditions under which a cell will respond to a pheromone or not. Under some conditions a cell does not respond at all. If a cell responds positively, there are two possible methods for its response: either the response is solely dependent on the initial concentration of its core component proteins in $\psi$ or the response is to some extent dependent on the concentration of the proteins in $\lambda$ as well. Experiment 2 looks for possible changes that the cell might adopt so that it can mate in conditions which made it respond negatively in Experiment 1. This is simulated by allowing the cell to utilize more concentration of proteins in $\lambda$. The results reveal that the cell can overcome the detrimental effects of the conditions by using more concentration of additional proteins in $\lambda$. These 2 experiments provide evidence that employing more concentration of proteins might be one of the ways that the cell uses to adapt itself in inhibiting conditions to facilitate mating. Experiment 3 tries to look for specific proteins in $\lambda$ that might be responsible for allowing a cell to change its response
to pheromone from positive to negative. The results reveal that in some case the protein set $\sigma$ is sufficient in regulating the response of the cell. In other cases, the requirements for the proteins in $\sigma$ are more stringent. Results of Experiments 4, 5 and 6 reveal that there are certain conditions (edge weights) in the model that are more important in determining whether a cell will respond positively or not.

8.2 Future Work

As a follow up of this work, we would like probe more about the functionality of the proteins in set $\lambda$. In Experiment 3 we looked at the performance of a subset of proteins ($\sigma$) in $\lambda$. We plan to extend our simulation to individual proteins in the set $\sigma$. This can be done by isolating a particular protein and varying its available concentration in the simulations. There is also scope of improving the model on several aspects. In our experiments the conditions that the cell is subjected to, in the simulation i.e., the edge weights used in the model are generated randomly. We can add some constraints on how weights are assigned to these edges, so that simulation can mirror real-world phenomena more closely. Also, in our model the number of tokens exchanged during interaction of places and transitions are integers as ordinary Petri net allows only that. However, in real life, the $kd$ value of reactions can hardly be expected to be integers all the time. So we would like to modify our model so that it can handle the exchange of fractional tokens among its nodes. Now in the pheromone pathway, evidence of negative feedback loop has been found, which has not been implemented in our model. So we would also like to explore some other variant of Petri net which allows implementation of such negative feedback loops. Finally we would like to extend our work to other unicellular organisms apart from yeast, to study their pheromone pathways and try to identify possible similarities
between the pheromone pathway across species.
Bibliography


[9] Dr. Steven Harris, personal communication.

[11] FEI LIU, MONIKA HEINER, and CHRISTIAN ROHR, *The manual for colored petri nets in snoopy - qpn\textsuperscript{c}/spn\textsuperscript{c}/cpn\textsuperscript{c}/ghpn\textsuperscript{c}* , Tech. report, Faculty of Mathematics, Natural Sciences and Computer Science, Brandenburg University of Technology, Cottbus, March 2012.


