Identification of cell signaling pathways based on biochemical reaction kinetics repositories

Student: Gustavo Estrela

Advisor: Marcelo da Silva Reis (Butantan Institute)

May 2019

Instituto de Matemática e Estatística
Centro de Toxinas, Resposta-imune e Sinalização Celular (CeTICS)
Laboratório Especial de Ciclo Celular, Instituto Butantan

This project receives funding from FAPESP

Cell Signaling

Cell signaling allows cells to respond to signals that come from its environment changing its behaviour accordingly.

Cell Signaling

Cell signaling allows cells to respond to signals that come from its environment changing its behaviour accordingly.

This mechanism is essential for many cell functions, including reproduction, growth and death.

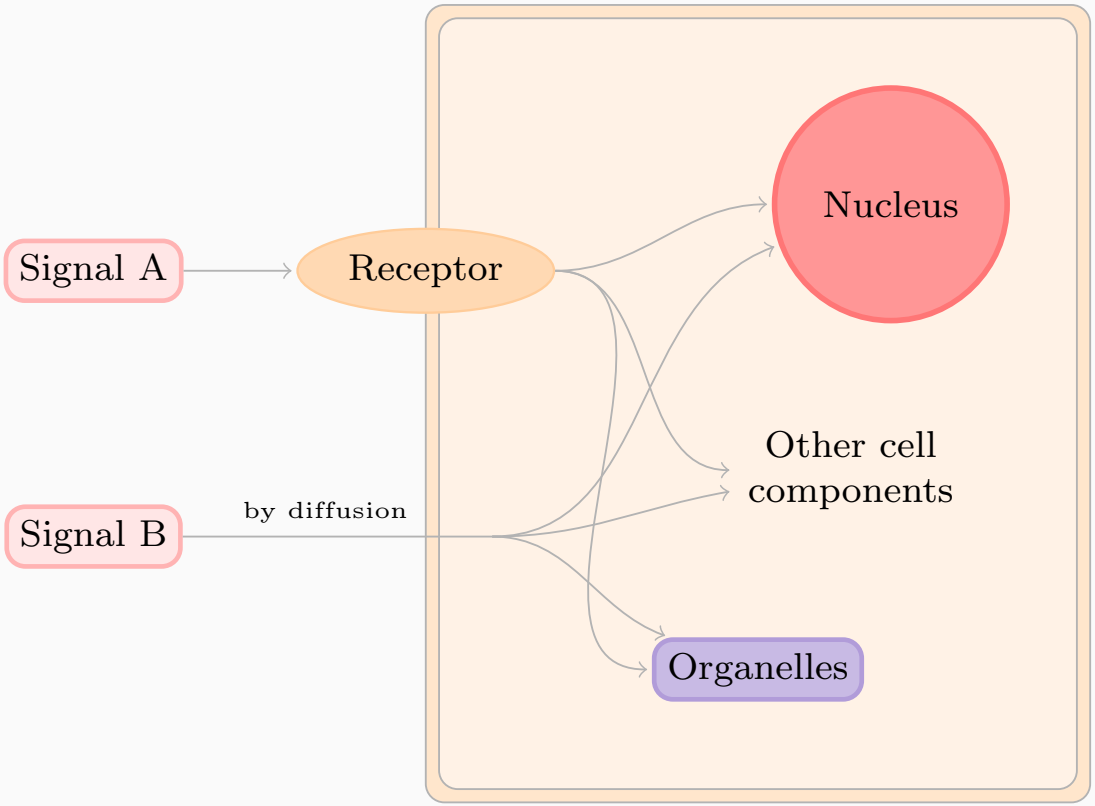
Cell Signaling

Cell signaling allows cells to respond to signals that come from its environment changing its behaviour accordingly.

This mechanism is essential for many cell functions, including reproduction, growth and death.

Understanding the functioning of cell signaling is important in many biological areas.

Cell Signaling



Cell Signaling

A signal propagates in an cell through chemical reactions that are caused by the change of concentration of chemical species.

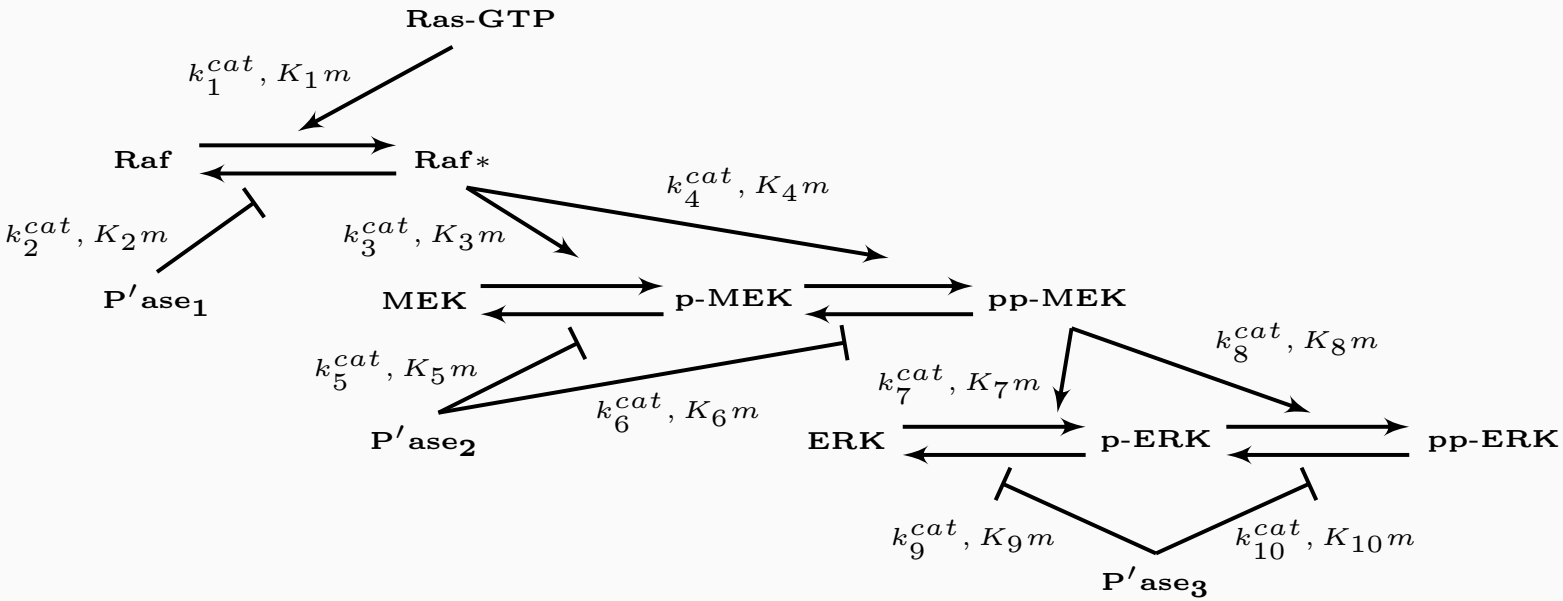
Cell Signaling

A signal propagates in an cell through chemical reactions that are caused by the change of concentration of chemical species.

We call the path of a signal a cell signaling pathway.

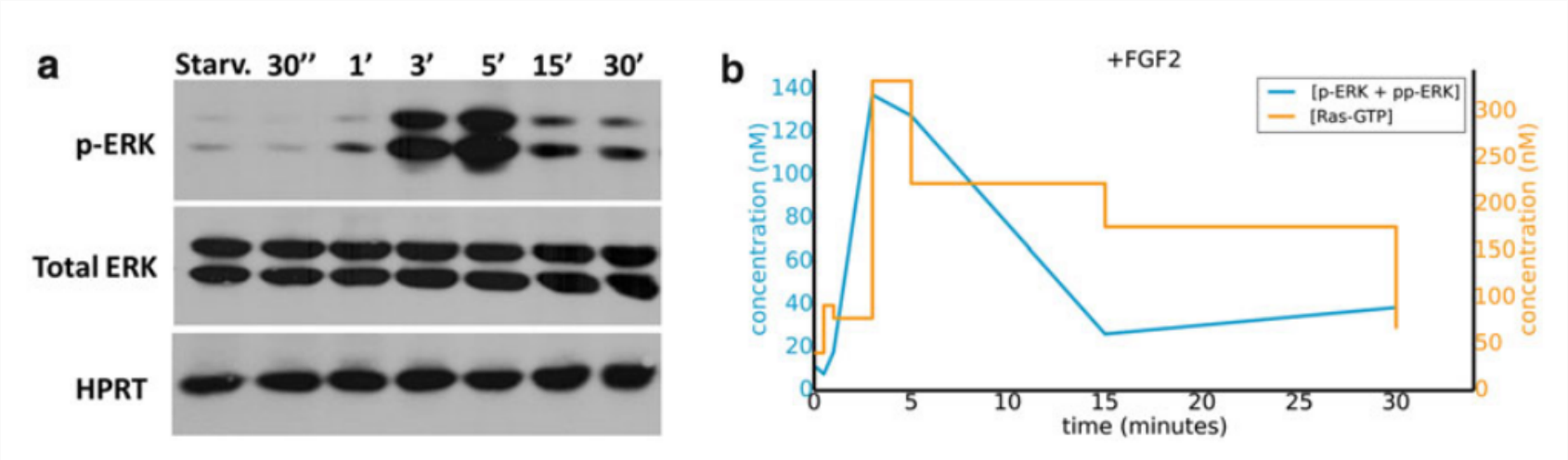
Cell Signaling Pathways

A cell signaling network can be characterized by a sequence of chemical reactions



Dynamic of Cell Signaling Pathways

We can summarize the state of the cell signaling pathway with measurements based on the concentration of some chemical species.



Identification of Cell Signaling Pathways

The problem of identification of cell signaling pathways is the problem of finding the components of a signaling pathway and how they interact given a set of experimental measurement.

Identification of Cell Signaling Pathways

The problem of identification of cell signaling pathways is the problem of finding the components of a signaling pathway and how they interact given a set of experimental measurement.

As the input, a description of a biological experiment and a set of experimental measurements are given.

Identification of Cell Signaling Pathways

The problem of identification of cell signaling pathways is the problem of finding the components of a signaling pathway and how they interact given a set of experimental measurement.

As the input, a description of a biological experiment and a set of experimental measurements are given.

A possible output to the problem is composed by:

Identification of Cell Signaling Pathways

The problem of identification of cell signaling pathways is the problem of finding the components of a signaling pathway and how they interact given a set of experimental measurement.

As the input, a description of a biological experiment and a set of experimental measurements are given.

A possible output to the problem is composed by:

- a model composed by a set of chemical reactions that are relevant for the biological experiment;

Identification of Cell Signaling Pathways

The problem of identification of cell signaling pathways is the problem of finding the components of a signaling pathway and how they interact given a set of experimental measurement.

As the input, a description of a biological experiment and a set of experimental measurements are given.

A possible output to the problem is composed by:

- a model composed by a set of chemical reactions that are relevant for the biological experiment;
- information about the reaction rate constants of the model.

A Model for Cell Signaling Pathways

We model cell signaling pathways with a system of ordinary differential equations. These equations represent the dynamic of concentrations of chemical species.

As an example, for the chemical reaction

$$\mathrm{A} + \mathrm{B} \longrightarrow \mathrm{C},$$

As an example, for the chemical reaction

$$\mathrm{A} + \mathrm{B} \longrightarrow \mathrm{C},$$

we can write the following equation:

$$\frac{d[\mathrm{C}]}{dt} = k[\mathrm{A}][\mathrm{B}];$$

where k is a reaction rate constant.

Identification of Cell Signaling Pathways

How can one find which reactions are relevant for a given biological experiment? One can find good candidates by:

Identification of Cell Signaling Pathways

How can one find which reactions are relevant for a given biological experiment? One can find good candidates by:

- using a priori information about the experiment;

Identification of Cell Signaling Pathways

How can one find which reactions are relevant for a given biological experiment? One can find good candidates by:

- using a priori information about the experiment;
- looking at biological databases such as KEGG.

Identification of Cell Signaling Pathways

How can one find which reactions are relevant for a given biological experiment? One can find good candidates by:

- using a priori information about the experiment;
- looking at biological databases such as KEGG.

However, that might produce incomplete models, or models with unimportant reactions.

Identification of Cell Signaling Pathways

How can one find which reactions are relevant for a given biological experiment? One can find good candidates by:

- using a priori information about the experiment;
- looking at biological databases such as KEGG.

However, that might produce incomplete models, or models with unimportant reactions.

It is then desirable to systematically propose new models.

Identification of Cell Signaling Pathways

Lulu Wu (2015) presented in her master dissertation a methodology that proposes to systematically modify models of signaling network in order to better represent experiments.

Identification of Cell Signaling Pathways

Lulu Wu (2015) presented in her master dissertation a methodology that proposes to systematically modify models of signaling network in order to better represent experiments.

On her work, the problem of identification of cell signaling pathways is treated as a feature selection problem.

Feature Selection Problem

The feature selection problem can be defined in the following way:

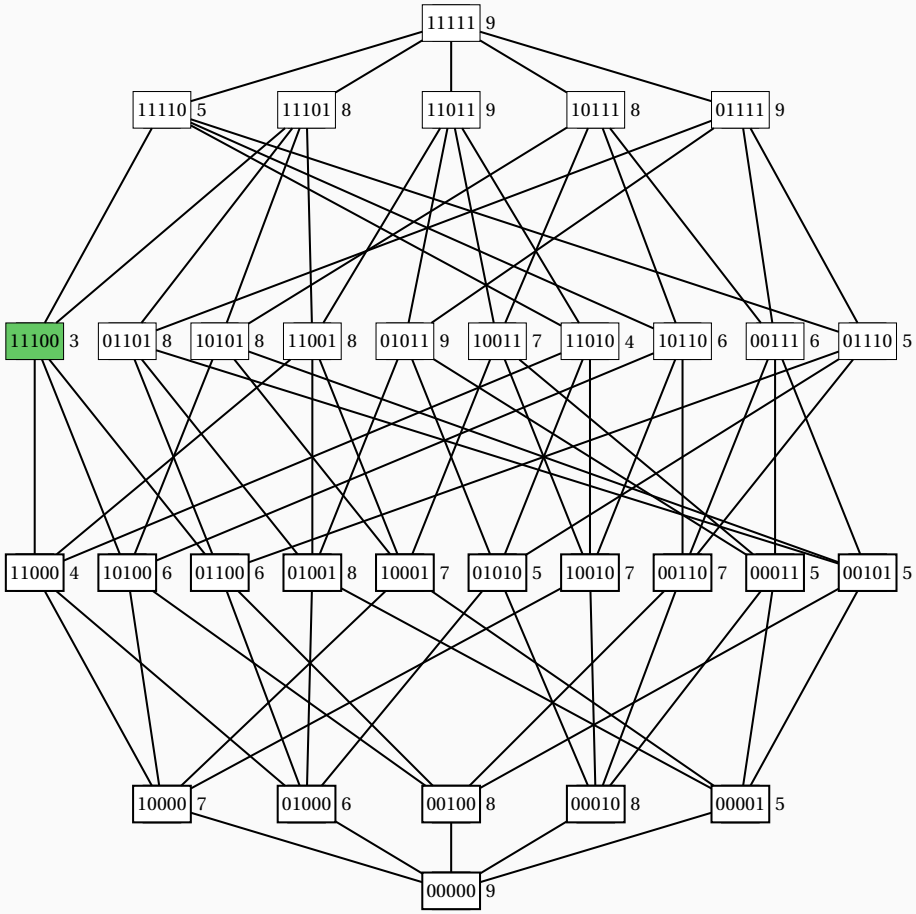
Given a set of features S and a cost function c, find subset $X \in \mathcal{P}(S)$, with minimum cost $c(X)$.

Feature Selection Problem

The feature selection problem can be defined in the following way:

Given a set of features S and a cost function c, find subset $X \in \mathcal{P}(S)$, with minimum cost $c(X)$.

Feature Selection Problem



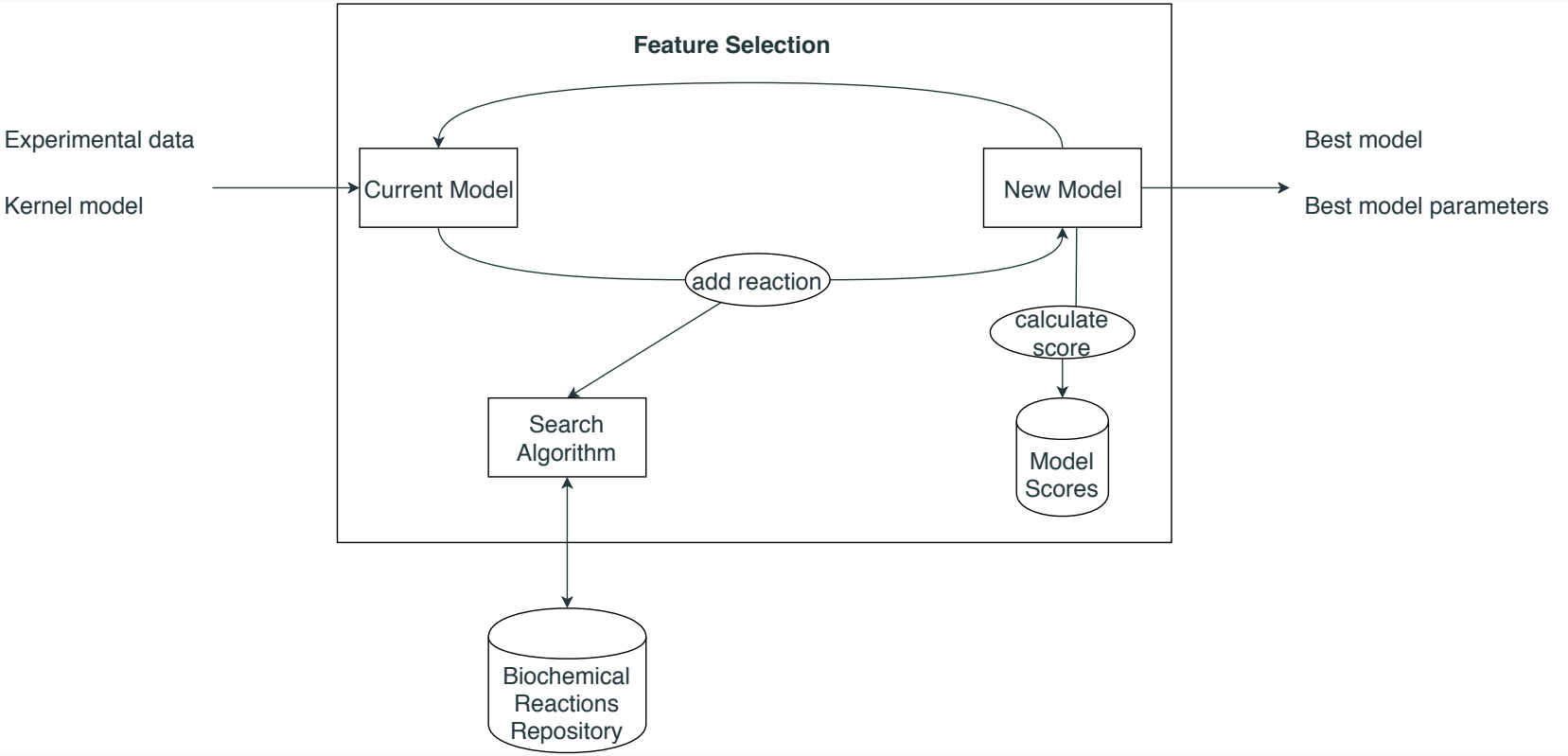
Feature Selection for Identification of Signaling Pathways

The methodology proposed by Wu defines the set of features as a set of chemical reactions that can be added to a starting model.

Feature Selection for Identification of Signaling Pathways

The methodology proposed by Wu defines the set of features as a set of chemical reactions that can be added to a starting model.This set of chemical reactions is fetched from KEGG and stored in a database of interactions.

Dynamic of Cell Signaling Pathways



Results of Wu's Methodology

However, the methodology worked satisfactorily only when the Kernel model was similar to the original model.

Difficulties of Wu's Methodology

We can point three aspects of Wu's work that could explain its limitations.

Difficulties of Wu's Methodology

We can point three aspects of Wu's work that could explain its limitations.

- the database of interactions used could be more nearly complete;

Difficulties of Wu's Methodology

We can point three aspects of Wu's work that could explain its limitations.

- the database of interactions used could be more nearly complete;
- the search algorithm could also consider removing interactions;

Difficulties of Wu's Methodology

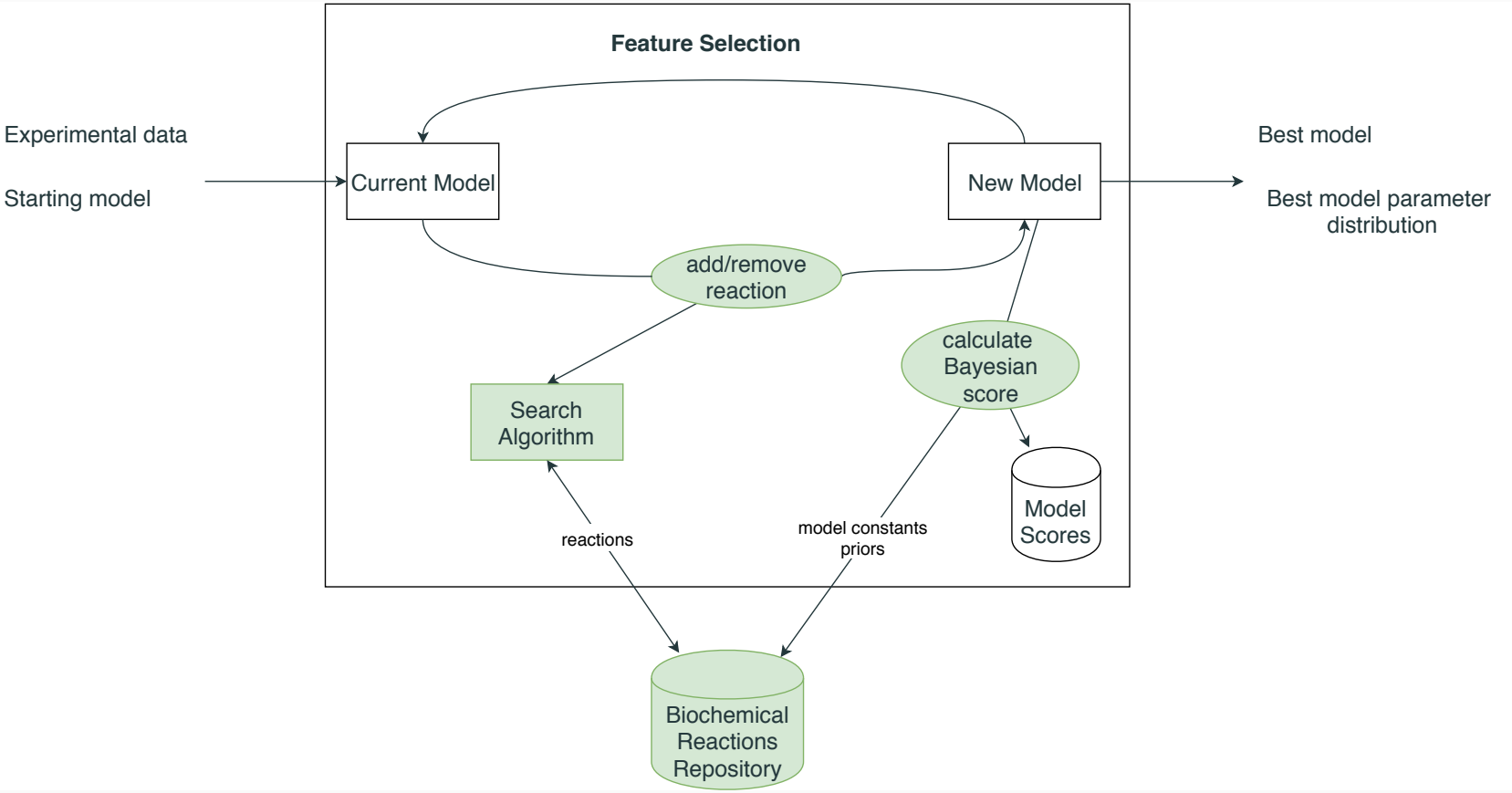
We can point three aspects of Wu's work that could explain its limitations.

- the database of interactions used could be more nearly complete;
- the search algorithm could also consider removing interactions;
- the cost function could implement a proper penalization of models;

What we Propose on this Project

We propose to create a methodology that uses a feature selection approach for identification of signaling pathways, tackling the difficulties encountered by Wu.

What we Propose on this Project



What we Propose on this Project

To get a more nearly complete database of interactions, we should fetch information from KEGG and other databases,

What we Propose on this Project

To get a more nearly complete database of interactions, we should fetch information from KEGG and other databases, such as STRING and SABIO-RK.

What we Propose on this Project

To use new search algorithms,

What we Propose on this Project

To use new search algorithms, we intend to use more general algorithms that can also remove interactions.

What we Propose on this Project

To define new cost functions,

What we Propose on this Project

To define new cost functions, we intend to use Bayesian approaches of model selection that allow us to create estimates of probabilities such as $p(M|D)$ or $p(D|M)$.

Objectives of this Project

Objectives of this Project

- Build a database of interactions.

Objectives of this Project

- Build a database of interactions.
- Create a cost function for models of signaling pathways.

Objectives of this Project

- Build a database of interactions.
- Create a cost function for models of signaling pathways.
- Formulate systematic modifications to a model as the search space of a feature selection model.

Objectives of this Project

- Build a database of interactions.
- Create a cost function for models of signaling pathways.
- Formulate systematic modifications to a model as the search space of a feature selection model.
- Test the methodology on known signaling pathways and also on pathways of interest in our lab.

Bayesian Ranking of Models

SigNetMS

To perform model ranking, we created a Python package called SigNetMS.

SigNetMS

To perform model ranking, we created a Python package called SigNetMS.

Given a experimental data, a model and its constants priors, the software can calculate an estimative of $p(D|M)$. This value is called the marginal likelihood.

SigNetMS

SigNetMS is an open source software and it is available at:

https://github.com/gustavoem/SigNetMS

Experiments on Model Selection

Model Ranking Experiment

We tested SigNetMS on the ranking of models.

Experiment description

The experiment is based on the following procedure:

Experiment description

The experiment is based on the following procedure:

- Create 4 candidate models.

Experiment description

The experiment is based on the following procedure:

- Create 4 candidate models.
- For one of the models, choose a set of parameter values and time steps and simulate data.

Experiment description

The experiment is based on the following procedure:

- Create 4 candidate models.
- For one of the models, choose a set of parameter values and time steps and simulate data.
- Add Gaussian noise to the simulations. Repeat two more times to generate three observations of the system.

Experiment description

The experiment is based on the following procedure:

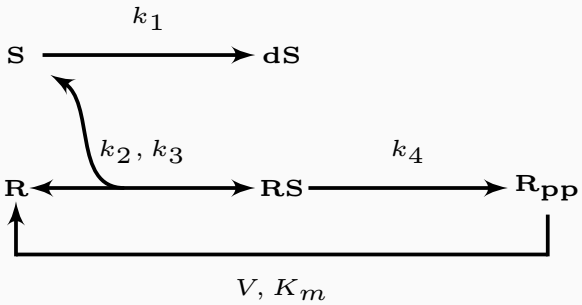
- Create 4 candidate models.
- For one of the models, choose a set of parameter values and time steps and simulate data.
- Add Gaussian noise to the simulations. Repeat two more times to generate three observations of the system.
- Neglect chosen parameter values and define prior distributions for every parameter.

Experiment description

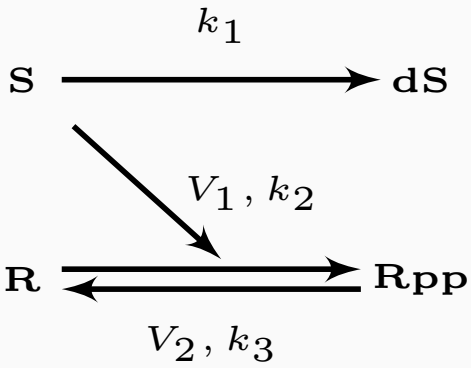
The experiment is based on the following procedure:

- Create 4 candidate models.
- For one of the models, choose a set of parameter values and time steps and simulate data.
- Add Gaussian noise to the simulations. Repeat two more times to generate three observations of the system.
- Neglect chosen parameter values and define prior distributions for every parameter.
- Rank the four models.

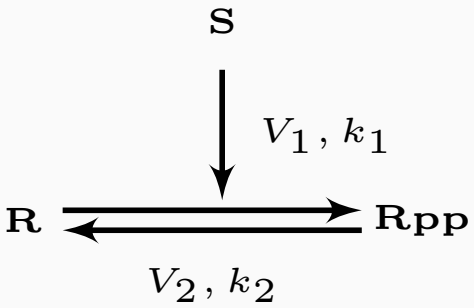
Model Ranking Experiment



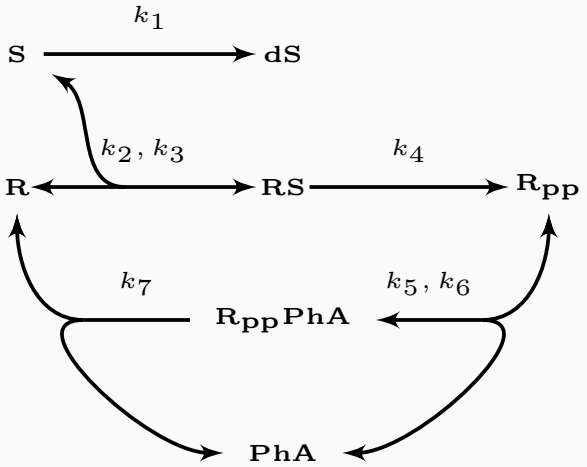
Model 1



Model 2



Model 3



Model 4

Results on SigNetMS

The ranking returned by SigNetMS on the experiment is:

$$1 > 2 > 4 > 3;$$

Results on SigNetMS

The ranking returned by SigNetMS on the experiment is:

$$1 > 2 > 4 > 3;$$

Note that:

Results on SigNetMS

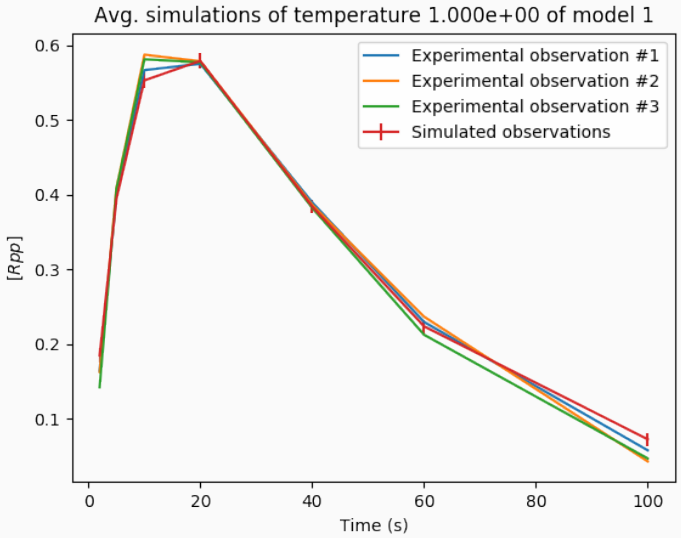
The ranking returned by SigNetMS on the experiment is:

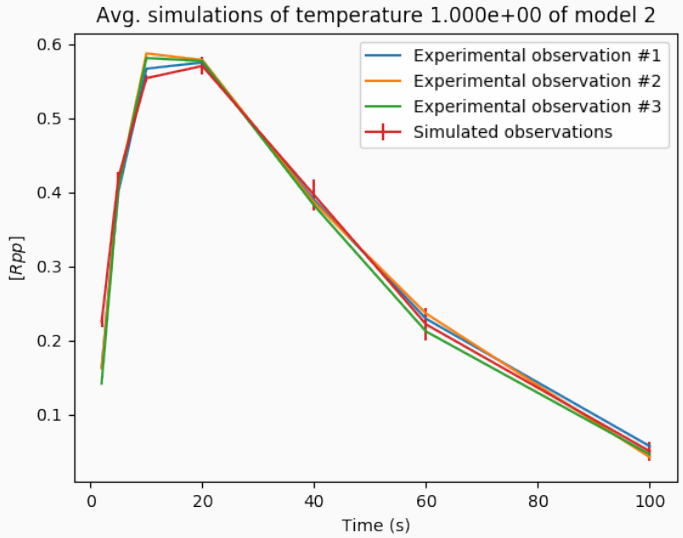
$$1 > 2 > 4 > 3;$$

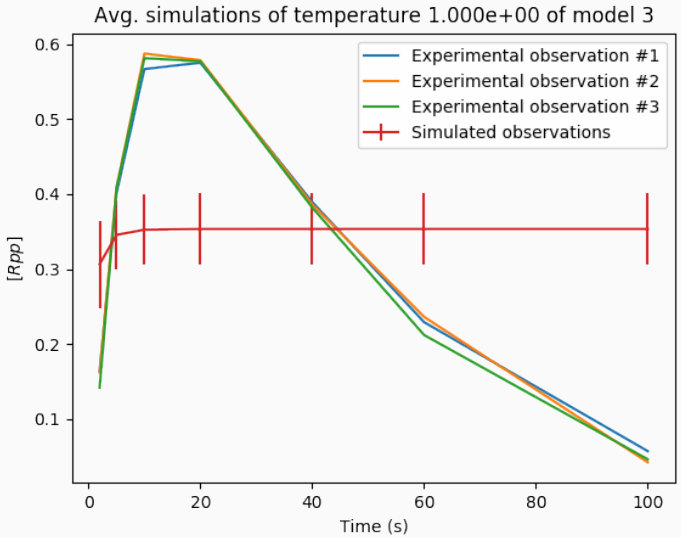
Note that:

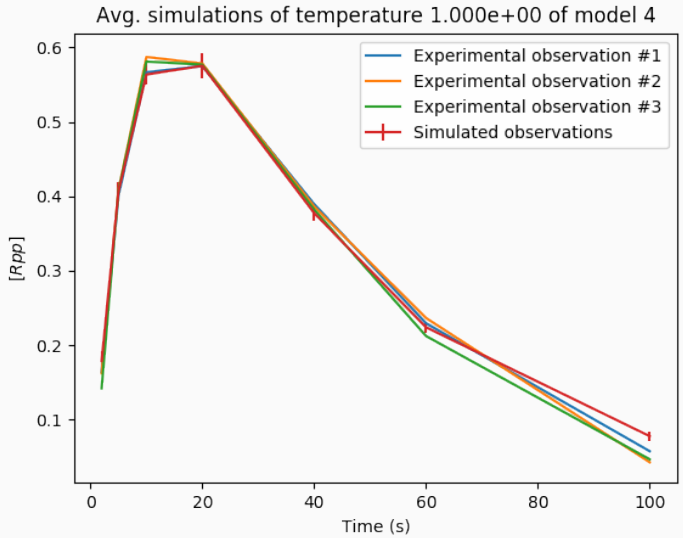
- the correct model is ranked first;
- overly complex models are ranked worse than simpler models.

Results on SigNetMS









Surface Walk Experiment

We also designed an experiment to walk on search space of models. The experiment consists in starting with a base model and then iteratively add reactions to it randomly, calculating their scores.

Surface Walk Experiment

We also designed an experiment to walk on search space of models. The experiment consists in starting with a base model and then iteratively add reactions to it randomly, calculating their scores.

The scores are calculated in respect to experimental data generated by a model that is present in the search space.

Surface Walk Experiment

the starting model

Surface Walk Experiment

the base model

Surface Walk Experiment

gráfico com valores calculados na cadeia percorrida
*provavelmente um comentário sobre a penalização dos modelos, e sobre o fato de que a cadeia percorrida não passa pelo modelo correto, o que fez os custos apenas subissem. *

Future Work

Future Work

More activities are expected to be completed in this project, mainly the follow:

- Embedding SigNetMS as a cost function for featsel, a framework for feature selection.

Future Work

More activities are expected to be completed in this project, mainly the follow:

- Embedding SigNetMS as a cost function for featsel, a framework for feature selection.
- Creation of a relational database of chemical interactions focused on our further applications.

Future Work

- Choice of a feature selection search algorithm.

Future Work

- Choice of a feature selection search algorithm.
- Apply the method in ERK signaling pathways of tumor cell lines Y1 and HEK293.

Thank you!