Stochastic Modeling and Analysis of Biological Networks

Ashish Tiwari

Tiwari@csl.sri.com

Computer Science Laboratory

SRI International

Menlo Park CA 94025

http://www.csl.sri.com/~tiwari

Collaborators: Carolyn Talcott, Merrill Knapp, Patrick Lincoln, Keith Laderoute

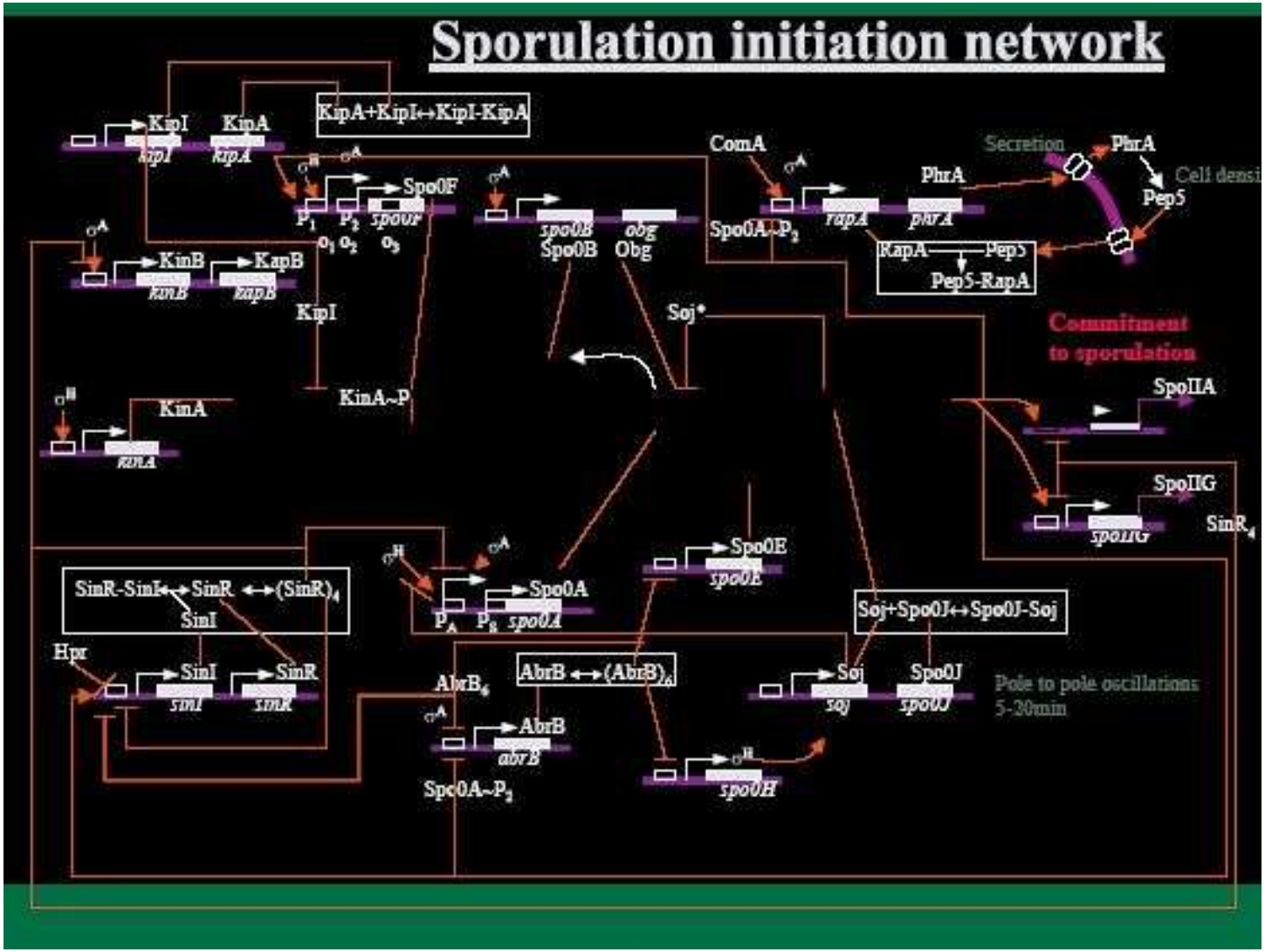
Introduction

- Biological data and models are large
- Meta-data on biological knowledge is huge
- When we have all the information required, for say risk assessment, how will we process this exponentially large information?
- Need efficient scalable algorithmic techniques to help us

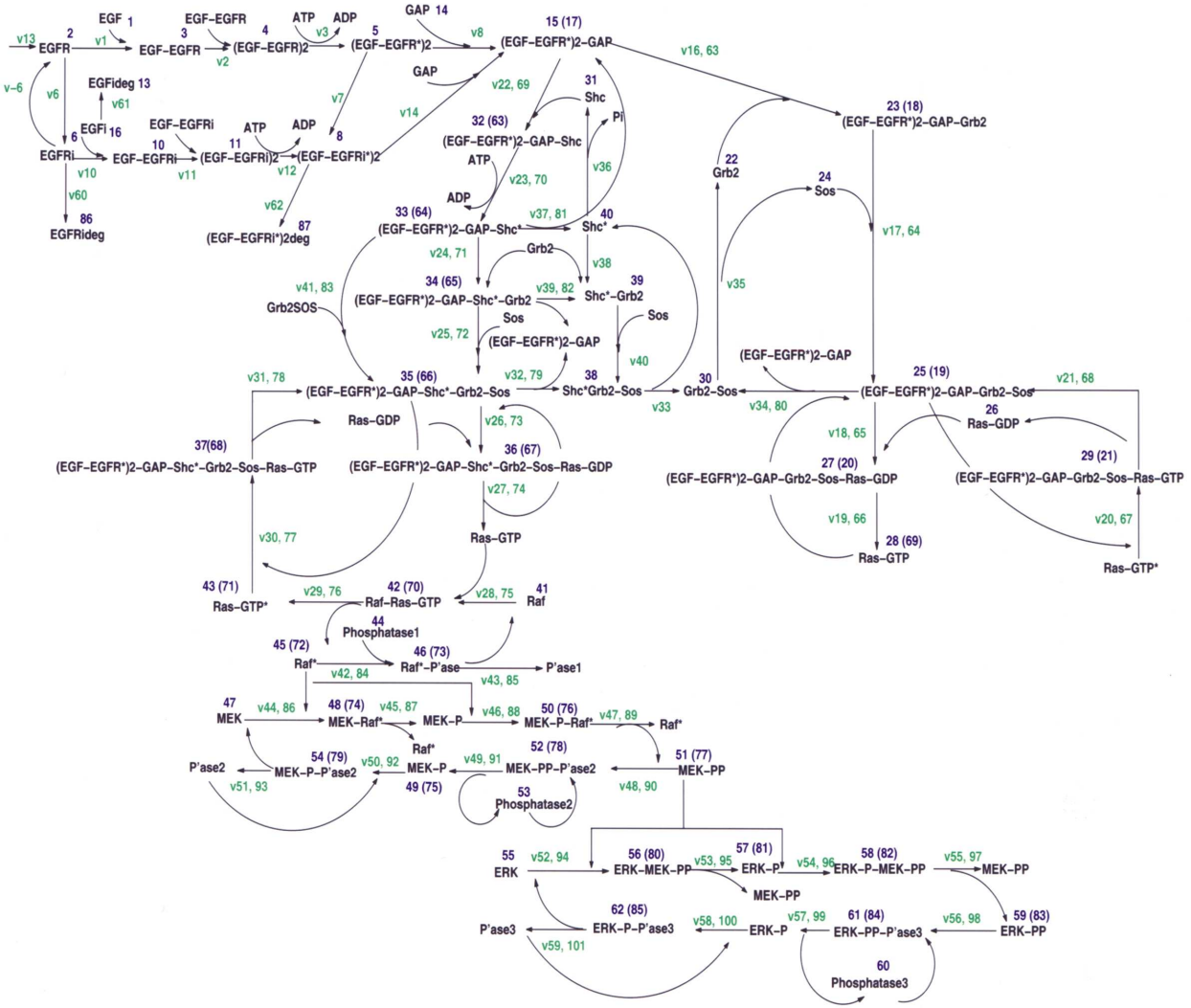
Representing Information: Reaction Networks

- Biological processes are often described as a collection of "reactions"
- Signaling pathways, metabolic pathways, regulatory pathways, . . ., internet
- Building a full kinetic model requires filling in the several unknown parameters, such as the reaction rates
- Goal: Analyze networks without complete specification of all its parameters, just based on its qualitative structure

Sporulation Initiation in *B.Subtilis*



EGF induced Erk Activation Pathway



Generic Reaction Network

Species S:

- molecule, ion, protein, enzyme, ligand, receptor, complex, modified form of protein
- web pages, threat sources, situational descriptors, events

Reactions R:

$$s_1, s_2 \xrightarrow{m_1, m_2} p_1, p_2$$

$$reactants \xrightarrow{modifiers} products$$

Anything that minimalistically captures the dynamics over the species

Traditional Kinetic Model

Ordinary differential equations extracted from the reaction network

Large number of unknown parameters

Parameters estimated so as to fit experimental data

Often low faith in the values of parameters and the model thus obtained

Goal and Approach

Goal: Analyze generic reaction networks, without complete specification of all its parameters, just based on its qualitative structure

Approach: Two novel ideas –

1. Define a notion of a **RANK** – based on a Markovian interpretation of reaction networks – of each species;
   Compute rank of each species using fast algorithms
2. Use the dual model – where reactions are the state variables and compute steady-states on the dual model

Stochastic Petrinet Semantics

For each species s_i, X_i denotes the number of molecules of s_i

State-space: $\vec{X} = [X_1, \ldots, X_n]$ is a n-dimensional vector of natural numbers

A reaction network defines a Markov process over this state space:

- From a state $\vec{X}$, one of the reaction $r_j \in R$ fires with probability $Pr(r_j \mid \vec{X})$

$$\vec{X} \quad \overset{Pr(r_j|\vec{X})}{\longmapsto} \quad \vec{X} + \vec{\nu}_j$$

where the probability is given by

$$Pr(r_j \mid \vec{X}) \quad = \quad \frac{1}{\alpha(\vec{X})} prop(r_j \mid \vec{X})$$

The Chemical Master Equation

Assuming that

$prop(r_j \mid \vec{X})dt$: the probability that, in the state $\vec{X}$, reaction r_j will occur once, somewhere inside the fixed volume, in the next infinitesimal time interval $[t, t + dt)$.

Time evolution of $P(\vec{X}, t \mid \vec{X}_0, t_0)$ is

$$\frac{\partial}{\partial t} P(\vec{X}, t \mid \vec{X}_0, t_0) = \sum_{r_j \in R} P(\vec{X} - \vec{v}_j, t \mid \vec{X}_0, t_0) prop(r_j \mid \vec{X} - \vec{v}_j) - prop(r_j \mid \vec{X}) P(\vec{X}, t \mid \vec{X}_0, t_0)$$

Our Markov process is the time abstract version.

Space-Partitioning Based Analysis

Y_i : probability that there is one molecule of species s_i in some small volume

Given $\vec{Y}(t)$, we can compute $\vec{Y}(t+1)$ as follows:

$$Y_i(t+1) = \sum_{r_j : s_i \notin (P \cup R)(r_j)} Pr(r_j \mid \vec{Y}(t)) \times Y_i(t) + \sum_{r_j : s_i \in P(r_j)} Pr(r_j \mid \vec{Y}(t)) \times 1$$

Assuming homogeneity, $\vec{Y}$ provides a good estimate for $\vec{X}$

Pathway Rank

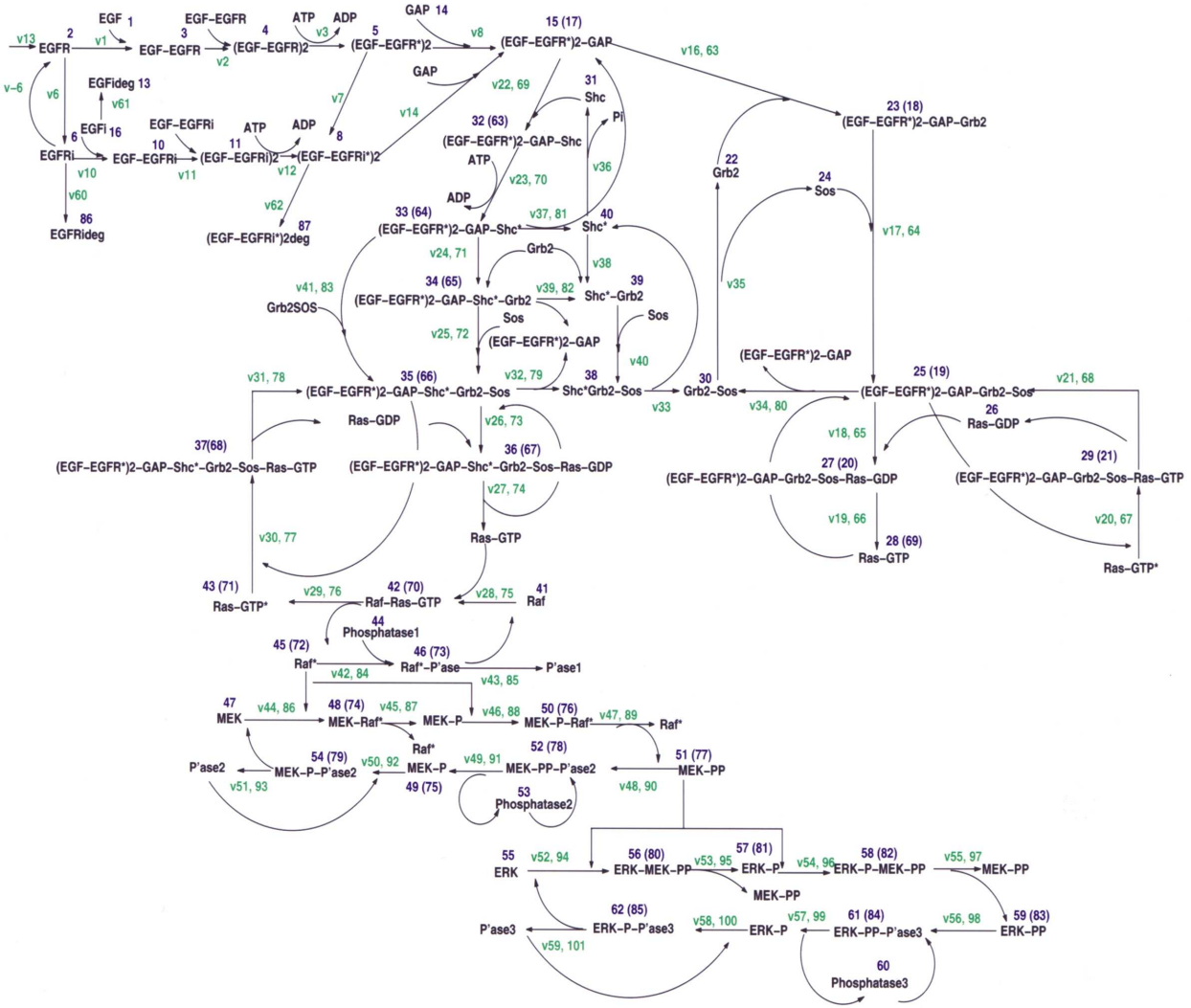
Starting with an initial probability distribition $\vec{Y}$, the analysis procedure attempts to compute the steady-state distribution

Can be understood as defining the rank of the species in reaction networks

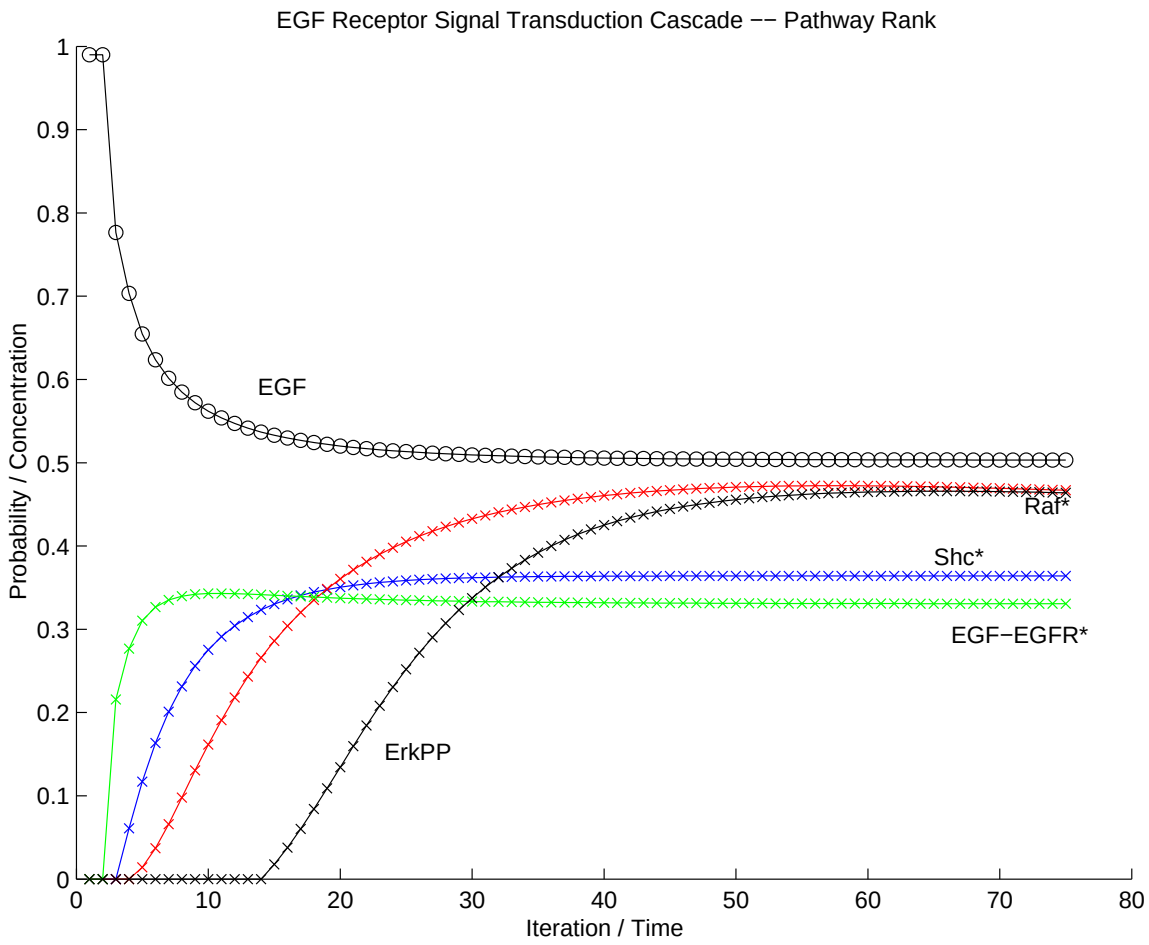
Advantages:

- System is never divergent for any choice of the propensity function; it is always stable or oscillatory
- Enzymatic reactions handled naturally; ODE approach requires tweaking
- Scalable approach

EGF Receptor Signal Transduction Cascade

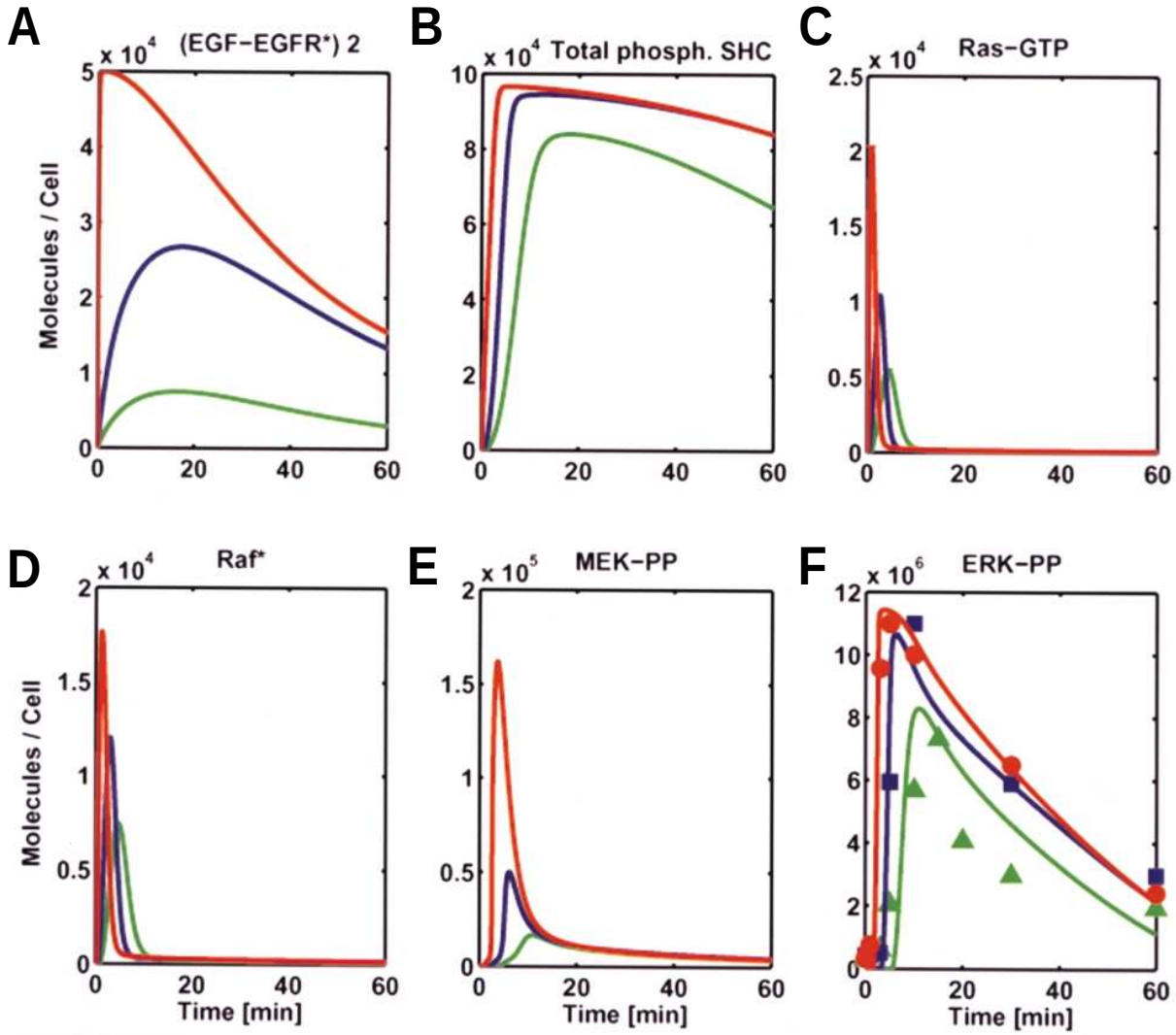


EGFR Signal Transduction: Results

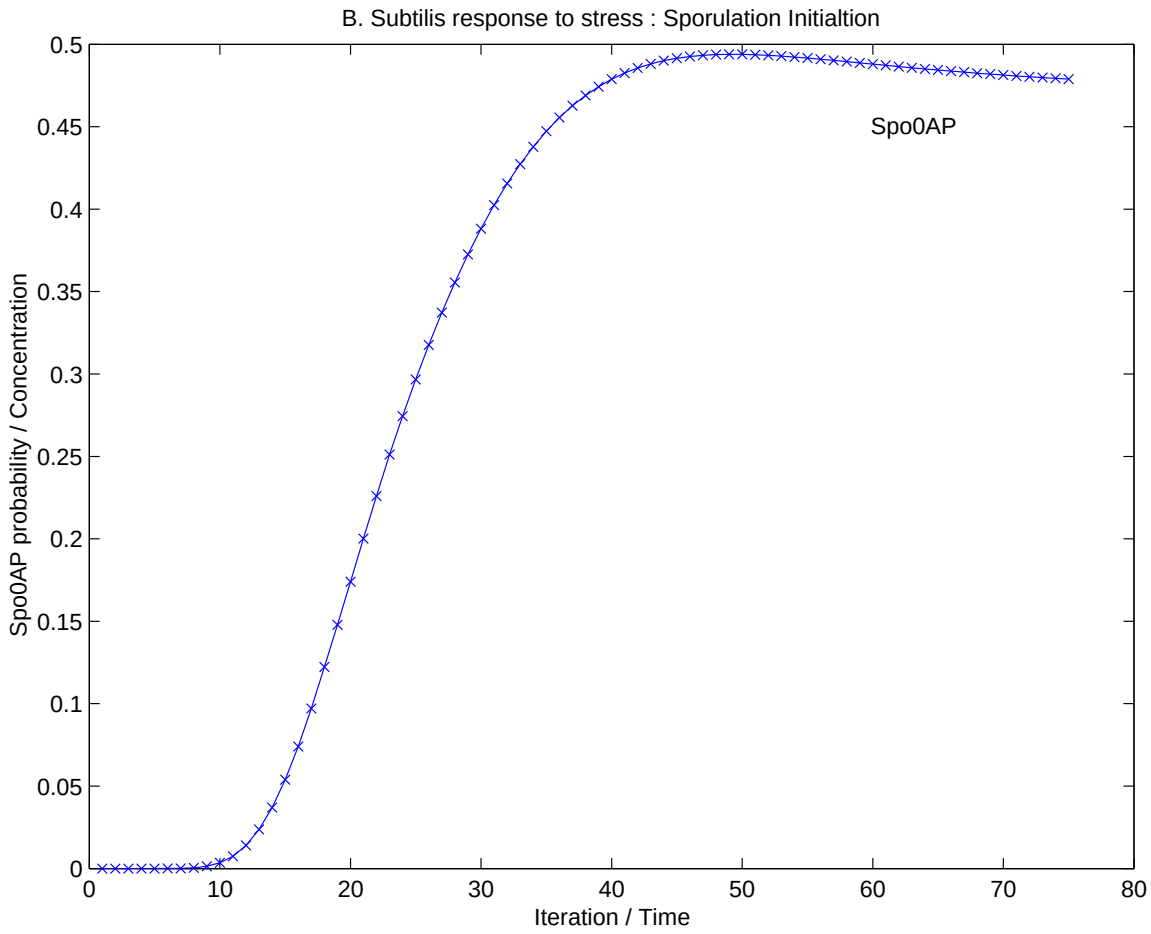


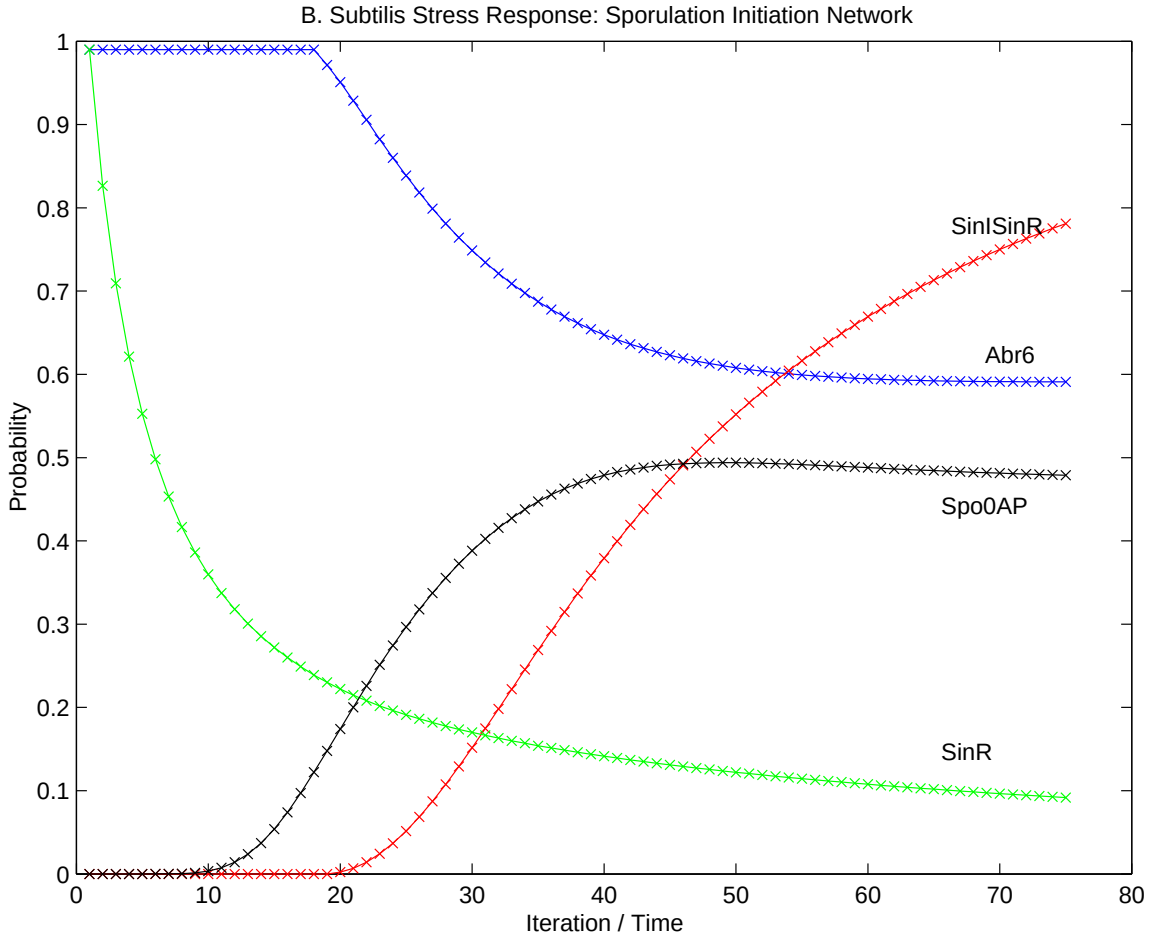
Using the same propensity function for all reactions

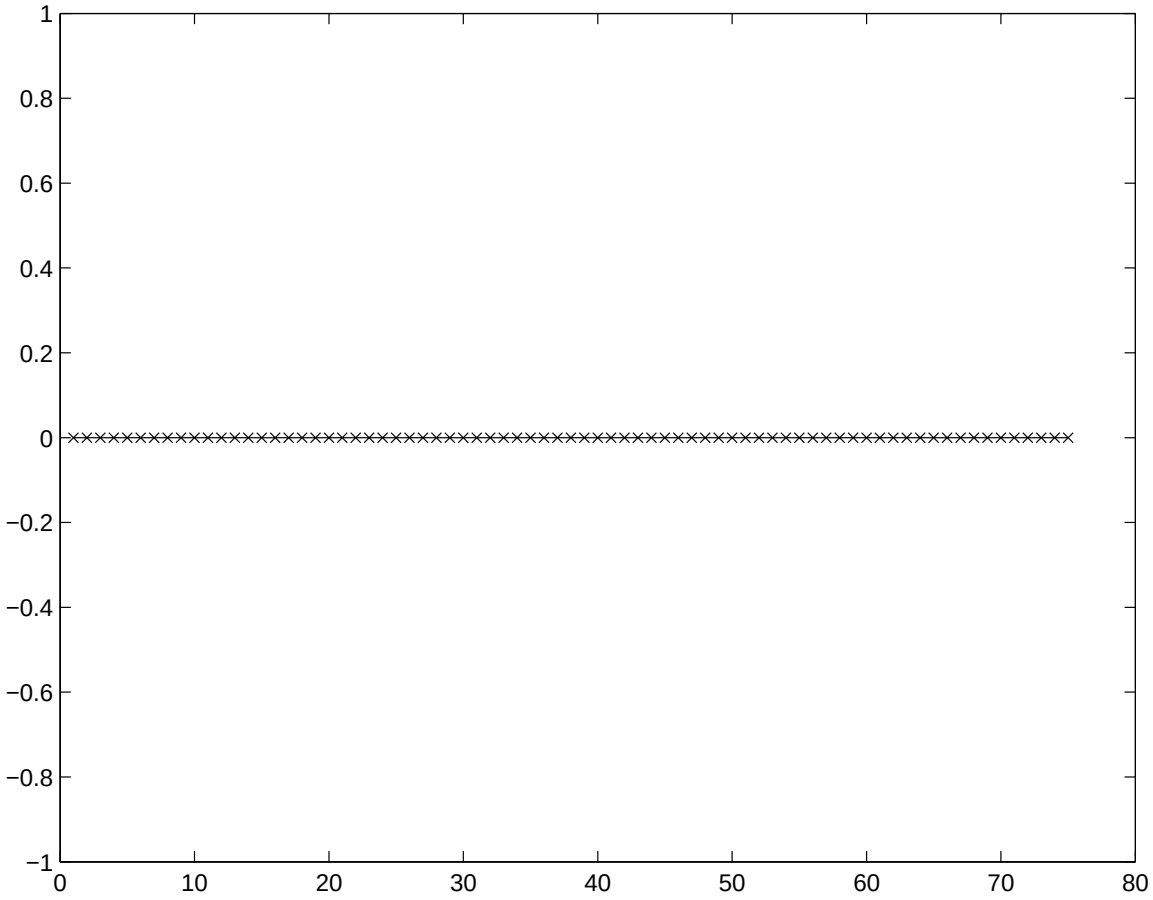
EGFR Signal Transduction: Kinetic Model



Sporulation Initiation in *B. Subtilis*




B. Subtilis Stress Response: Sporulation Initiation Network
Probability
Iteration / Time
SinISinR
Abr6
Spo0AP
SinR
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
10
20
30
40
50
60
70
80


1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1
0
10
20
30
40
50
60
70
80

Part II: The Dual Approach

Fast Analysis Using Boolean SAT Approach

Reaction Networks – A Dual Approach

Reactions, and not species, define the state space

A reaction can be on or off

The reaction network is interpreted using the two basic rules:

- if a reaction is “off”, but its reactants and modifiers are present, then the reaction is turned “on”
- if a reaction is “on”, but one of its reactants or modifiers is not present, then the reaction is turned “off”

A species is present if it is the product of some “on” reaction and not the reactant of any “on” reaction

Reaction Networks To Boolean SAT

The steady-state in this model is a
set of reactions that can be consistently on

Steady-state configurations can be efficiently detected using modern SAT solvers

Specific / desired steady-state configurations can be detected using weighted MaxSAT solvers

EGF Stimulation Network

Being developed in Pathway Logic Project

Model of EGF stimulation by curating reactions involved in mammalian cell signaling

For model validation,

- Started with 400 reactions
- Added initial species in the dish
- Specified a set of target species that are experimentally observed in response to EGF stimulation

EGF Stimulation Network: Results

Analysis results:

- No solution without violating a competitive inhibition constraint in the MaxSAT instance
- Several syntactic errors in the model detected and corrected
- (Frap1:Lst8)-CLc identified as the conflict causing species
- This leads to two hypotheses
  - (Frap1:Lst8)-CLc splits into two *populations* one for each of the two competing reactions;
  - there is a feedback loop that can reset the state of (Frap1:Lst8)-CLc and the system oscillates between the two pathways.

  Experiments are ongoing to test these hypotheses.

MAPK Signaling Network

Mitogen-Activated Protein kinase (MAPK) network regulates several cellular processes, including the cell cycle machinery

Model from BhallaRamIyeger, Science 2002 and BhallaIyenger, Chaos 2001

Analysis finds two stable sets of behavior:

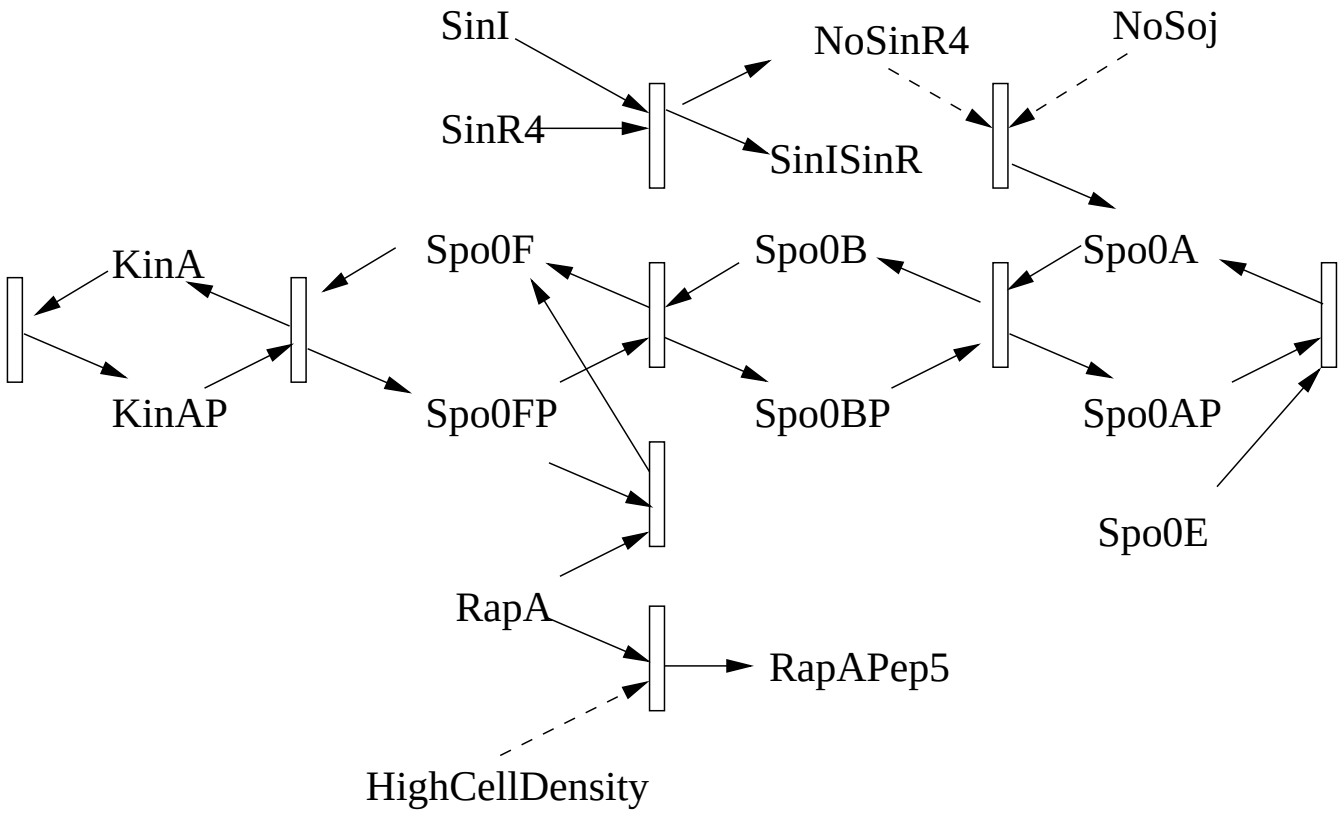
- The positive feedback loop is active:

  $$Grb2, Sos1, PKC* \mapsto Ras \mapsto Raf* \mapsto Mek* \mapsto Erk* \mapsto AA* \mapsto PKC*$$

- The negative feedback loops are active: $PP2A$ dephosphorylates both Raf* and Mek*, and MKP dephosphorylates Erk*. MKP is created by transcription of *MKP* gene, and this is promoted by Erk*.

Overall system behavior is a result of the interaction between the positive and negative cycles.

Sporulation Initiation in *B. Subtilis*



Analysis of Sporulation Initiation Network

The tool finds 3 different behaviors:

- sporulation initiated:
  - SinI produced
  - SinI binds to SinR
  - Preventing SinR from repressing *spo0A*
  - RapA converted to RapAPep5,
  - Preventing RapA from dephosphorylating Spo0A-P
  - Presence of stress signals prevent KipI from inhibiting KinA from self-kinasing
  - Self-kinasing of KinA triggers the phosphorelay
  - Leads to production of Spo0A-P

Analysis of Sporulation Initiation Network

- Not enough cell-density:
  - RapA dephosphorylates Spo0F-P
  - Breaking the phosphorelay chain
  - Resulting in no production of Spo0A-P.
- The third stable state scenario is similar to the first, except that Spo0E dephosphorylates the produced Spo0A-P, thus using up the produced Spo0A-P.

The three stable scenarios each make different assumptions about the environment.

Summary

- Generic reaction networks is used commonly to represent biological knowledge, and it can be used to represent meta-knowledge
- To get detailed kinetic models requires estimating the large number of unknown parameters
- We presented two scalable approaches for analyzing generic reaction networks using its structural information
- These can be used to qualitatively understand hypothesized models, even in the detailed parameter information

Thank You!

For publications, visit:

`http://www.csl.sri.com/~tiwari/`