Inferring parameters in genetic regulatory networks

Camilo La Rota[1] Fabien Tarissan[2] Leo Liberti[2]

[1]Complex Systems Institute (IXXI)
Ecole Normale Superieure - CNRS, Lyon, France

[2]LIX (Computer science laboratory)
Ecole Polytechnique, Palaiseau, France

ARS Workshop, LIX-Polytechnique, Palaiseau, October 31 2008

Outline

1. Introduction
   - Inverse Problems in Biological Complex Systems
   - Biological Context
2. Modelling the Biological Problem
   - Gene Expression, regions and tissues
   - Gene Interaction Network
   - Gene Regulatory Network models
3. GRN Inference
   - Modelling the inverse problem
     - Defining the GRN
     - Defining the inverse problem
   - Mathematical Programming Formulation
     - Definitions
     - Objective Function and Constraints
     - Objective Function and Constraints
     - Reformulation and linearization

Outline

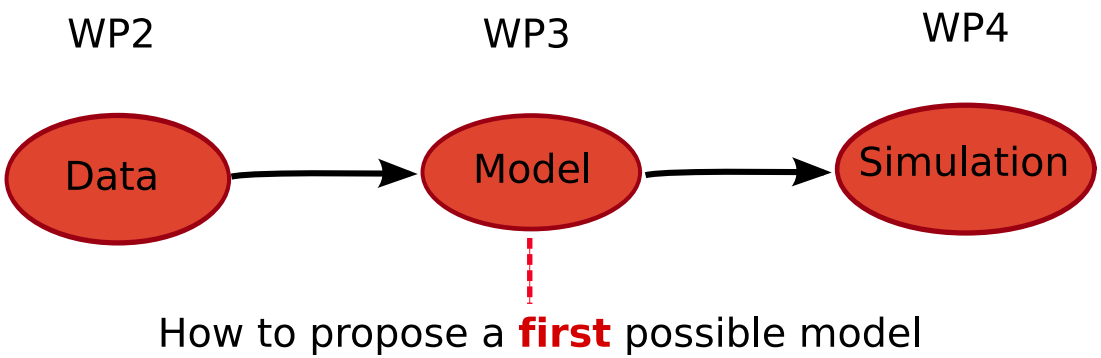
1. Introduction
   - Inverse Problems in Biological Complex Systems
   - Biological Context
2. Modelling the Biological Problem
   - Gene Expression, regions and tissues
   - Gene Interaction Network
   - Gene Regulatory Network models
3. GRN Inference
   - Modelling the inverse problem
     - Defining the GRN
     - Defining the inverse problem
   - Mathematical Programming Formulation
     - Definitions
     - Objective Function and Constraints
     - Objective Function and Constraints
     - Reformulation and linearization

European Morphex Project

Biological Problem Solving:
Gene regulatory networks and cell interactions in morphogenesis.
Models and protocols for parameter inference.

Complex Systems:
Meta-model and associated concepts for designing tools and protocols.

Simulation Platform:
Generic pre and post simulation tools and generic protocols.



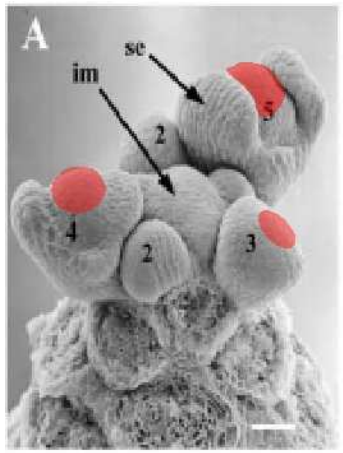
Outline

1. Introduction
   - Inverse Problems in Biological Complex Systems
   - Biological Context
2. Modelling the Biological Problem
   - Gene Expression, regions and tissues
   - Gene Interaction Network
   - Gene Regulatory Network models
3. GRN Inference
   - Modelling the inverse problem
     - Defining the GRN
     - Defining the inverse problem
   - Mathematical Programming Formulation
     - Definitions
     - Objective Function and Constraints
     - Objective Function and Constraints
     - Reformulation and linearization

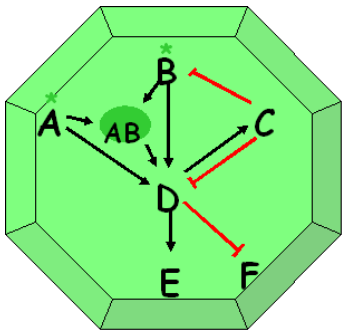
Genetic regulatory networks (GRN) and morphogenesis

Developmental stages of Arabidopsis Thaliana

- Arabidopsis Flower Development
  - GRN dynamics + other factors : morphogenesis, structure, tissue diversity
  - Continuous development
  - Discrete stages

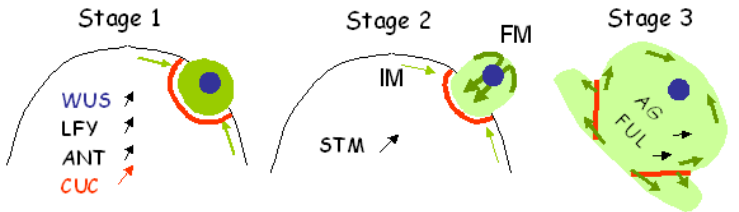


- Genetic Control of Morphogenesis



Recruitment of founder cells from the flanks of the IM.

FC proliferation, outgrowth begins.



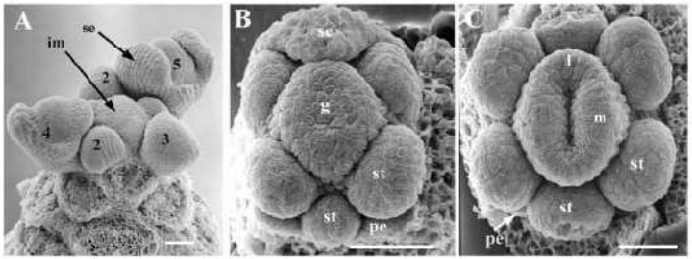
Establishement of Floral Meristem (FM). Sepal Primordia appears.

Establishement of carpel identity.

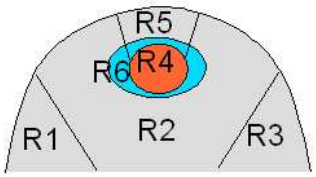
GRN Subnetworks' Stability

Mutants stable states $\sim$ Unstable states at wild-type stages.

Wild type stages (unstable states)

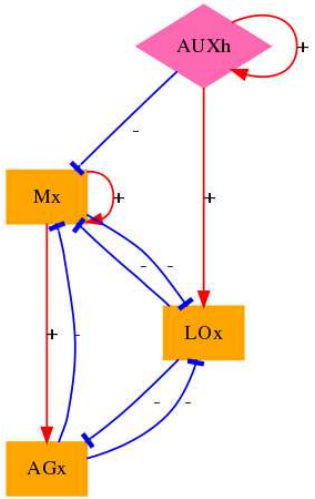


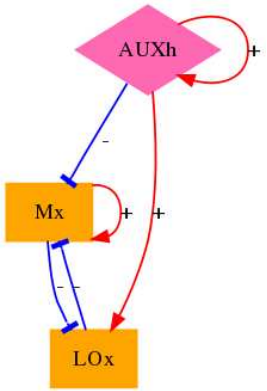
Unstable states stage 2 ~ (ag, pi) mutant stable states



(Pelaz, 2000)

stage 2 = at least 4 stable states (sepals (1) + meristem (3))





Outline

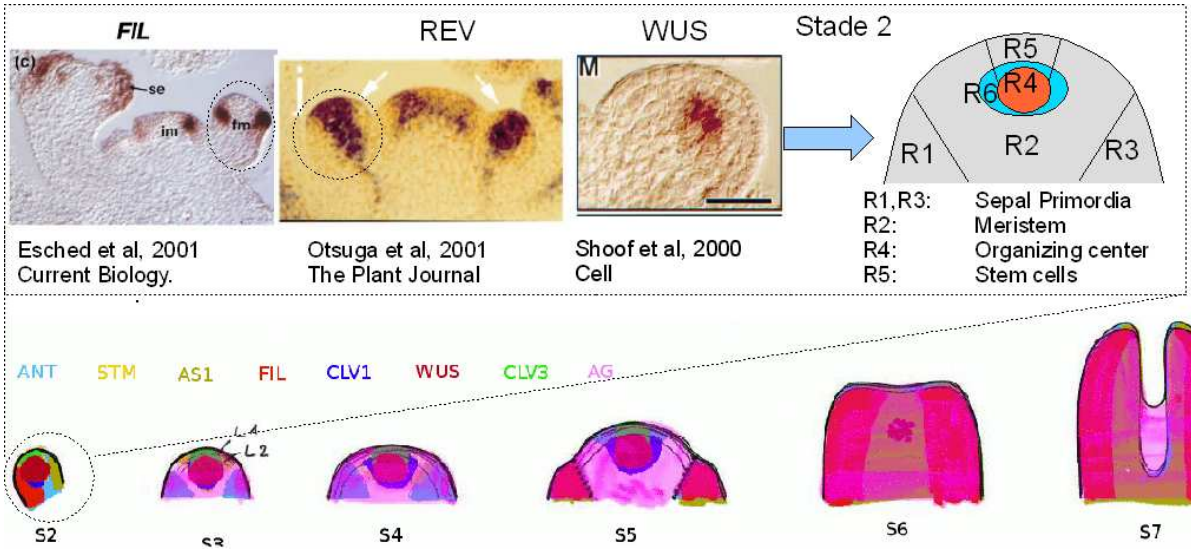
1. Introduction
   - Inverse Problems in Biological Complex Systems
   - Biological Context
2. Modelling the Biological Problem
   - Gene Expression, regions and tissues
   - Gene Interaction Network
   - Gene Regulatory Network models
3. GRN Inference
   - Modelling the inverse problem
     - Defining the GRN
     - Defining the inverse problem
   - Mathematical Programming Formulation
     - Definitions
     - Objective Function and Constraints
     - Objective Function and Constraints
     - Reformulation and linearization

Gene Expression, regions and tissues

Expression data

mRNA spatiotemporal distribution

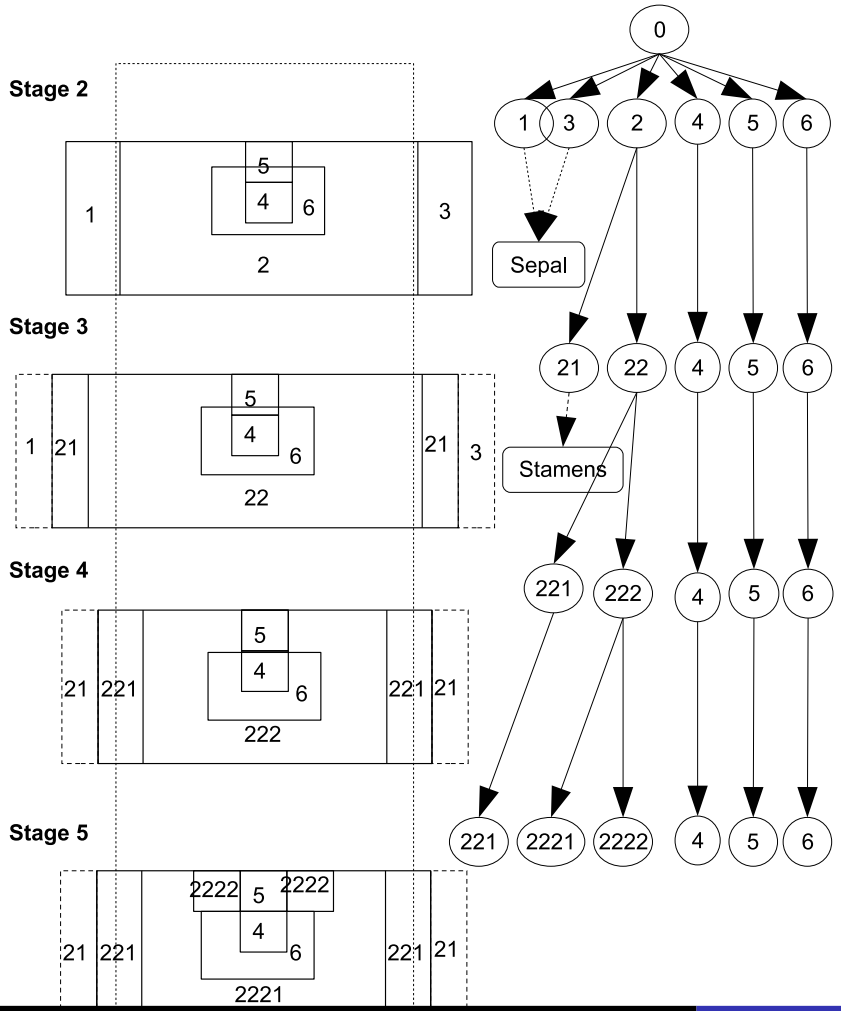
- Qualitative
- Imprecise
- Time-discrete

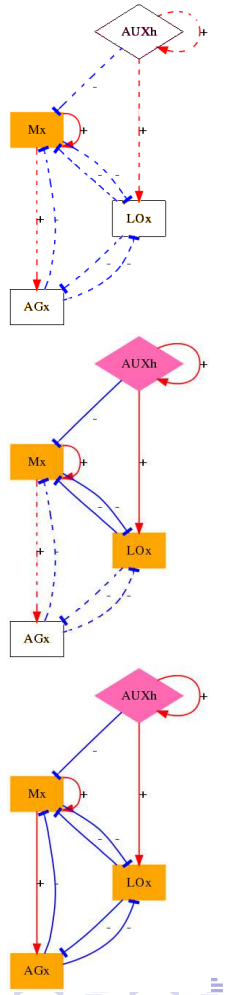


Exploiting the data

- Superposition of expression patterns reveals regions.
- Data is difficult to analyze, multiple interpretations are possible.
- Tentative subdivisions in homogeneous regions are proposed.

Cell or tissue lineage





Outline

1. Introduction
   - Inverse Problems in Biological Complex Systems
   - Biological Context
2. Modelling the Biological Problem
   - Gene Expression, regions and tissues
   - Gene Interaction Network
   - Gene Regulatory Network models
3. GRN Inference
   - Modelling the inverse problem
     - Defining the GRN
     - Defining the inverse problem
   - Mathematical Programming Formulation
     - Definitions
     - Objective Function and Constraints
     - Objective Function and Constraints
     - Reformulation and linearization

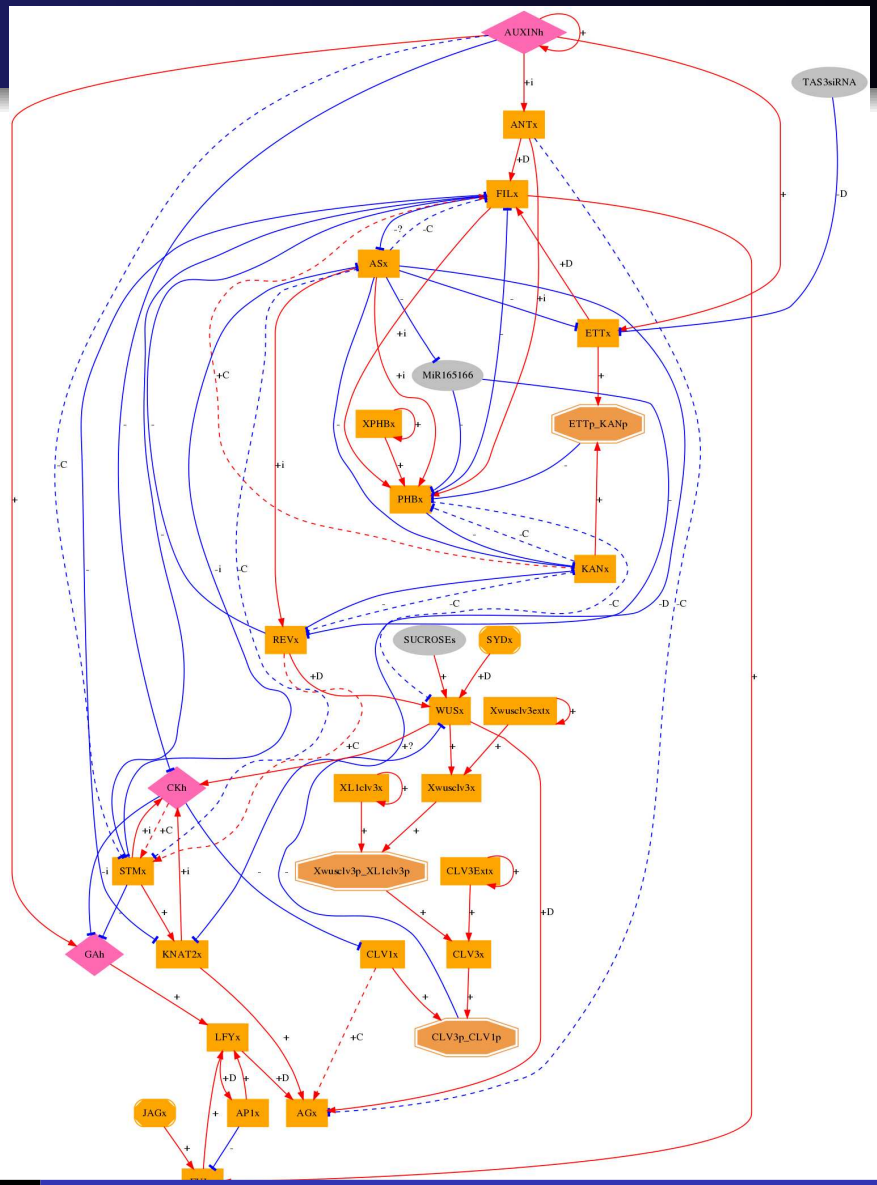
Gene Interaction Network

Interaction data

- Molecular evidence
- Genetic evidence

Exploiting the Data

- Uncertain
- Conflicting interpretations
- Error prone
- Prior Interaction Network



Outline

1. Introduction
   - Inverse Problems in Biological Complex Systems
   - Biological Context
2. Modelling the Biological Problem
   - Gene Expression, regions and tissues
   - Gene Interaction Network
   - Gene Regulatory Network models
3. GRN Inference
   - Modelling the inverse problem
     - Defining the GRN
     - Defining the inverse problem
   - Mathematical Programming Formulation
     - Definitions
     - Objective Function and Constraints
     - Objective Function and Constraints
     - Reformulation and linearization

Gene Regulatory Network models

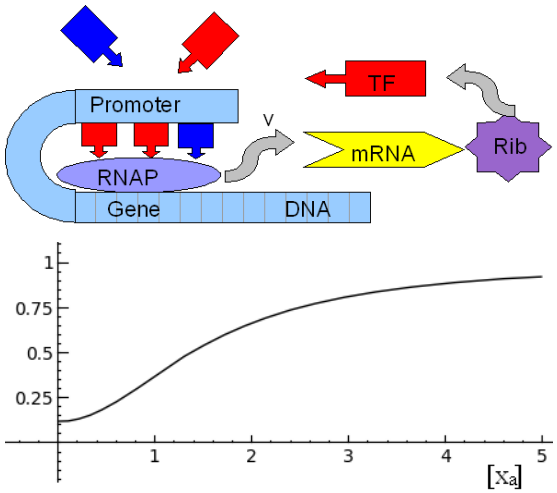
Gene transcription mechanisms, mass action kinetics: the Shea-Ackers model

Quantitative activity of gene i

$$\frac{d([x_i](t))}{dt} = f_i([P],[x_1],\ldots,[x_m]) - \lambda_i[x_i]$$

$$f_i([P],[x_1],\ldots,[x_m]) = \sum_{s\in S_i} \nu(s)\mathbb{P}(s_i = s)$$

$$\mathbb{P}(s_i = s) = \frac{K_B(s)[P]^{\alpha_s}[x_1]^{\alpha_s^1}\ldots[x_m]^{\alpha_s^m}}{1+\sum_{z\in S_i} K_B(z)[P]^{\alpha_z}[x_1]^{\alpha_z^1}\ldots[x_m]^{\alpha_z^m}}$$



Exemples of regulatory phenomena

Activation

$$f_i([P],[x_a]) =$$

$$\frac{[P](\nu_p K_p + \nu_{ap} K_{ap}[x_a])}{1+K_p[P]+K_a[x_a]+K_{ap}[x_a][P]}$$

Gene Regulatory Network models

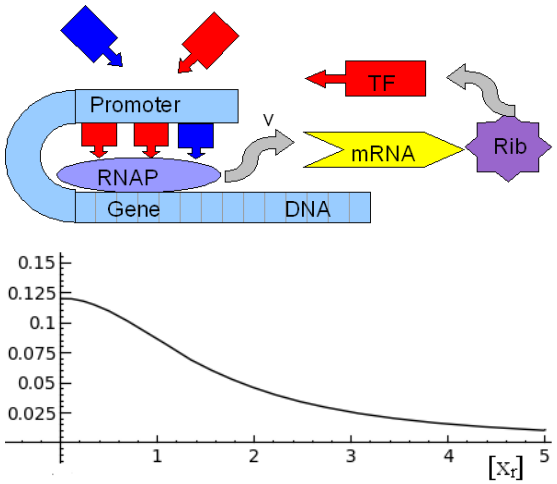
Gene transcription mechanisms, mass action kinetics: the Shea-Ackers model

Quantitative activity of gene i

$$\frac{d([x_i](t))}{dt} = f_i([P],[x_1],\ldots,[x_m]) - \lambda_i[x_i]$$

$$f_i([P],[x_1],\ldots,[x_m]) = \sum_{s\in S_i} \nu(s)\mathbb{P}(s_i = s)$$

$$\mathbb{P}(s_i = s) = \frac{K_B(s)[P]^{\alpha_s}[x_1]^{\alpha_s^1}\ldots[x_m]^{\alpha_s^m}}{1+\sum_{z\in S_i} K_B(z)[P]^{\alpha_z}[x_1]^{\alpha_z^1}\ldots[x_m]^{\alpha_z^m}}$$



Exemples of regulatory phenomena

Activation

$$f_i([P],[x_a]) = \frac{[P](\nu_p K_p + \nu_{ap} K_{ap}[x_a])}{1+K_p[P]+K_a[x_a]+K_{ap}[x_a][P]}$$

Repression

$$f_i([P],[x_r]) = \frac{[P]\nu_p K_p}{1+K_p[P]+K_r[x_r]+K_{rp}[x_r][P]}$$

Gene Regulatory Network models

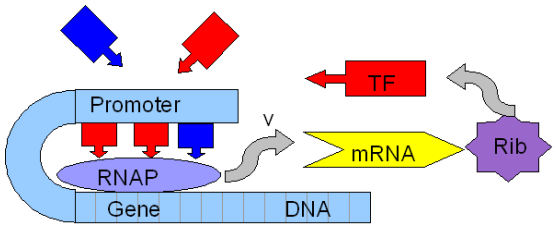
Gene transcription mechanisms, mass action kinetics: the Shea-Ackers model

Quantitative activity of gene i

$$\frac{d([x_i](t))}{dt} = f_i([P],[x_1],\ldots,[x_m]) - \lambda_i[x_i]$$

$$f_i([P],[x_1],\ldots,[x_m]) = \sum_{s\in S_i} \nu(s)\mathbb{P}(s_i = s)$$

$$\mathbb{P}(s_i = s) = \frac{K_B(s)[P]^{\alpha_s}[x_1]^{\alpha_s^1}\ldots[x_m]^{\alpha_s^m}}{1+\sum_{z\in S_i} K_B(z)[P]^{\alpha_z}[x_1]^{\alpha_z^1}\ldots[x_m]^{\alpha_z^m}}$$



Exemples of regulatory phenomena

Activation

$$f_i([P],[x_a]) = \frac{[P](\nu_p K_p + \nu_{ap} K_{ap}[x_a])}{1+K_p[P]+K_a[x_a]+K_{ap}[x_a][P]}$$

Repression

$$f_i([P],[x_r]) = \frac{[P]\nu_p K_p}{1+K_p[P]+K_r[x_r]+K_{rp}[x_r][P]}$$

Competition/Synergy

$$f_i([P],[x_1],\ldots,[x_m]) = \frac{[P](\nu_p K_p + \sum_{i\in\{1,\ldots,m\}} \nu_{ip} K_{ip}[x_i])}{1+K_p[P]+\sum K_i[x_i]+K_{ip}[x_i][P]}$$

Gene Regulatory Network models

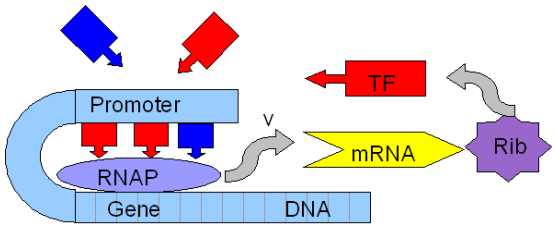
Quantitative Piecewise Differential and Qualitative Generalized Logical models

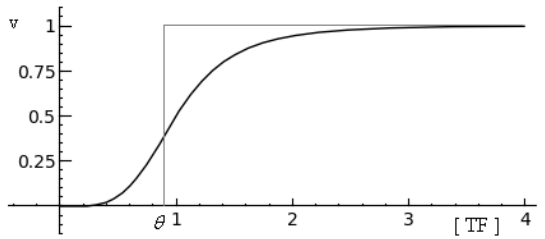
Quantitative activity of gene i

$$\frac{d(x_i(t))}{dt} = f_i(x_1, \ldots, x_m) - \lambda_i x_i(t)$$

$$f_i(\vec{x}(t)) = \sum_{j \in 1,\ldots,m} (\nu_{0i} + \nu_{ji} H^{\alpha_{ji}}(x_j(t), \sigma_{ji}))$$

$$x_i(t) = \frac{F_i(\vec{x}^0)}{\lambda_i} - \left(\frac{F_i(\vec{x}^0)}{\lambda_i} - x_i^0\right) e^{-\lambda_i t}$$





- σ_{ji}: threshold of interaction.
- ν_{ji}: induced transcription rate.
- α_{ij} : Kind of interaction (-1, +1)

Remarks

Lost: transitory dynamics, interaction crosstalk (constant thresholds).

Gene Regulatory Network models

Quantitative Piecewise Differential and Qualitative Generalized Logical models

Quantitative activity of gene i

$$\frac{d(x_i(t))}{dt} = f_i(x_1, \ldots, x_m) - \lambda_i x_i(t)$$
$$f_i(\vec{x}(t)) = \sum_{j \in 1,\ldots,m} (\nu_{0i} + \nu_{ji} H^{\alpha_{ji}}(x_j(t), \sigma_{ji}))$$
$$x_i(t) = \frac{F_i(\vec{x}^0)}{\lambda_i} - \left(\frac{F_i(\vec{x}^0)}{\lambda_i} - x_i^0\right) e^{-\lambda_i t}$$

- σ_{ji}: threshold of interaction.
- ν_{ji}: induced transcription rate.
- α_{ij} : Kind of interaction (-1, +1)

Qualitative activity of gene i

$$q_i(n) = \Delta(x_i(t_n), \{\sigma_{ji}\}_j)$$
$$\psi_i(n) = \Delta(f_i(\vec{x}(t_n))/\lambda_i, \{\sigma_{ji}\}_j)$$
$$\psi_i(n) = F_L(q_1(n), \ldots, q_m(n))$$
$$q_i(n+1) \rightarrow \psi_i(n)$$

- Δ: Discretization operator.
- ψ: Image of state $\vec{q}$.
- F_L : Multivalued function.

Remarks

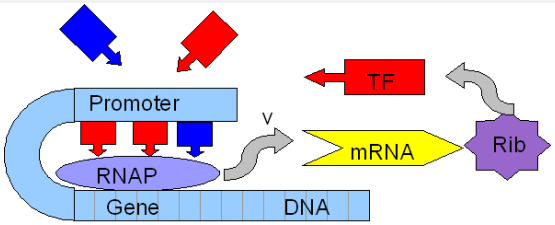
Lost: transitory dynamics, interaction crosstalk (constant thresholds).

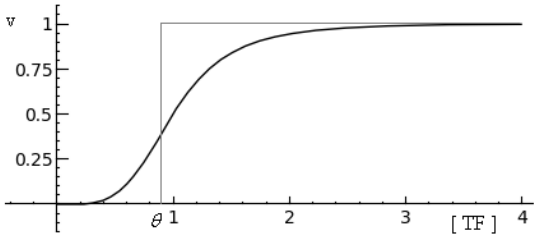
Gene Regulatory Network model

Weighted sum and threshold boolean network paradigm

Qualitative activity of gene i

$$q_i(n) = H(x_i(t_n), \sigma_i)$$
$$\psi_i(n) = H(\frac{f_i(\vec{x}(t_n))}{\lambda_i}, \sigma_i)$$
$$q_i(n+1) = \psi_i(n)$$
$$\frac{f_i(\vec{x}(t_n))}{\lambda_i} = \sum_{j \in 1,\ldots,m} (\frac{\nu_{0i}}{\lambda_i} + \frac{\nu_{ji}}{\lambda_i} H^{\alpha_{ji}}(x_j(t), \sigma_j))$$
$$q_i^{n+1} = H\left(\sum_{j=1}^{m} \alpha_{ij} w_{ij} q_j^n - \theta_i\right)$$





- θ_i: threshold of activation.
- w_{ij}: interaction strength $\left(\frac{(inducedproduction)}{decay}\right)$.
- α_{ij} : Kind of the interaction $(-1, +1)$

Remarks

Outline

1. Introduction
   - Inverse Problems in Biological Complex Systems
   - Biological Context
2. Modelling the Biological Problem
   - Gene Expression, regions and tissues
   - Gene Interaction Network
   - Gene Regulatory Network models
3. GRN Inference
   - Modelling the inverse problem
     - Defining the GRN
     - Defining the inverse problem
   - Mathematical Programming Formulation
     - Definitions
     - Objective Function and Constraints
     - Objective Function and Constraints
     - Reformulation and linearization

Modelling the inverse problem (I): defining the GRN

Solve steady state equations, no time evolution

Gene Regulatory Network (GRN): $(G, T, \alpha, w, x, \theta)$

- Sets and Graph:
  - V: vertexes (genes)
  - A: arcs (interactions)
  - $G = (V, A)$
- Functions:
  - $\alpha : A \to \{+1, -1\}$ *arc sign*;
  - $w : A \to \mathbb{R}_+$ *arc weight*;
  - $x : V \to \{0, 1\}$ *gene state*;
  - $y : V \to \{0, 1\}$ *state image*;
  - $\theta : V \to \mathbb{R}$ *threshold*,
- Evolution rules

$$y(v) = \begin{cases} 1 & \text{if } \sum_{u \in \delta^-(v)} \alpha(u,v) w(u,v) x(u) \geq \theta(v) \\ 0 & \text{otherwise,} \end{cases}$$

where $\delta^-(v) = \{u \in V \mid (u,v) \in A\}$ for all $v \in V$.

Modelling the inverse probleme: defining the problem

Finding network parameters for simultaneous stable subnetworks

Given

- (G, α)
- $S := \{1..Smax\}$: set of stages and/or mutants.
- $U = \{U_s\}_{s \in S}; U_s \subseteq V$: nodes of G_s, the (induced) subnetworks of G.
- $R = \{R_s\}_{s \in S}; R_s := \{1..Rmax_s\}$: regions of homogeneous expression.
- $\Phi = \{\phi_{s,r,u}\}_{s \in S, r \in R_s, u \in U_s}; \phi_{s,r,u} : V \rightarrow \{0,1\}$: expression data.

Find

w, θ such that all $(G_s, \alpha, w, \vec{x}_{s,r}, \theta)$ satisfy the steadiness constraints and collectively minimize the total $D_H(\vec{x}, \vec{\phi})$.

D_H : hamming distance from steady state (fixed point) to data.

Outline

1. Introduction
   - Inverse Problems in Biological Complex Systems
   - Biological Context
2. Modelling the Biological Problem
   - Gene Expression, regions and tissues
   - Gene Interaction Network
   - Gene Regulatory Network models
3. GRN Inference
   - Modelling the inverse problem
     - Defining the GRN
     - Defining the inverse problem
   - Mathematical Programming Formulation
     - Definitions
     - Objective Function and Constraints
     - Objective Function and Constraints
     - Reformulation and linearization

Mathematical Programming Formulation

$$\left.\begin{array}{rrcl} \min_x & f(x) & & \\ \text{subject to} & g(x) & \leq & 0, \end{array}\right\}$$

x: decision variables, f: objective function, g : constraints

Sets	V, A, S, R (genes, interactions, stages, regions)
Variables	$x : V \times R \rightarrow \{0,1\}$, $w : A \rightarrow \mathbb{R}^+$, $\theta : A \rightarrow \mathbb{R}$
Parameters	$\alpha : A \rightarrow \{-1,+1\}$, bounds: $\theta^L, \theta^U, w^L, w^U$
	$\phi_{v,r}$, (observed gene expression.)

Objective Function and Constraints

Objective function

$$\sum_{s \in S, r \in R_s} \sum_{u \in U_s} |x_{s,u,r} - \rho_{s,u,r}|$$

State image rules

$$\sum_{u \in U_s : (u,v) \in A} \alpha_{u,v} w_{u,v} x_{s,r,u} \geq \theta_v y_{s,r,v} - \|V\| (1 - y_{s,r,v})$$

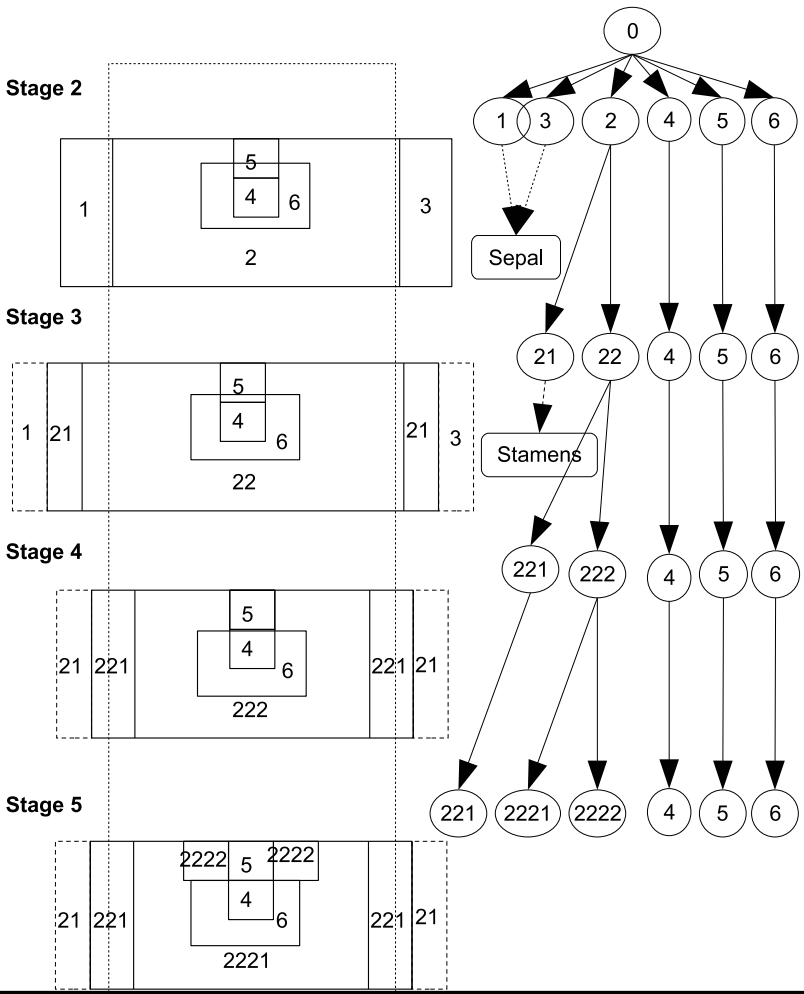
$$\sum_{u \in U_s : (u,v) \in A} \alpha_{u,v} w_{u,v} x_{s,r,u} \leq (\theta_v - \varepsilon)(1 - y_{s,r,v}) + \|V\| y_{s,r,v}$$

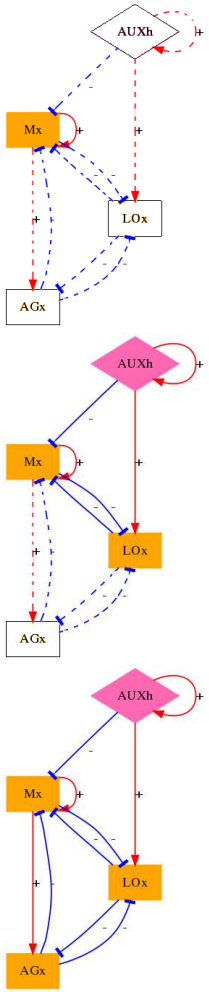
Steadiness conditions

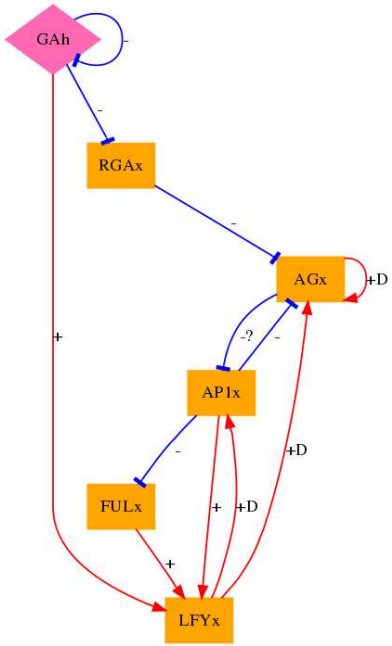
$$\forall s \in S, r \in R_s, u \in U_s \qquad y_{s,u,r} = x_{s,u,r}$$

Cell or tissue lineage:

Knowledge on steady states AND initial conditions







- Which transitory dynamics ?
- Which update scheduling ?

Modelling the inverse probleme(II): defining the GRN

Find fixed points from initial conditions

Gene Regulatory Network (GRN): $(G, T, \alpha, w, x, \iota, \theta)$

- Sets and Graph:
  - V: vertexes (genes)
  - A: arcs (interactions)
  - $T := \{1, 2, ..\} \subset \mathbb{N}$
  - $G = (V, A)$
- Evolution rules

- Functions:

$\alpha : A \to \{+1, -1\}$	*arc sign*;
$w : A \to \mathbb{R}_+$	*arc weight*;
$x : V \times T \to \{0, 1\}$	*gene activation*;
$\iota : V \to \{0, 1\}$	*initial configuration*;
$\theta : V \to \mathbb{R}$	*threshold*,

$$x(v,1) = \iota(v)$$

$$x(v,t) = \begin{cases} 1 & \text{if } \sum_{u \in \delta^-(v)} \alpha(u,v) w(u,v) x(u,t-1) \geq \theta(v) \\ 0 & \text{otherwise,} \end{cases}$$

where $\delta^-(v) = \{u \in V \mid (u,v) \in A\}$ for all $v \in V$

Modelling the inverse problem: defining the problem

Finding network parameters for simultaneous stable subnetworks,using initial condition data

Given

- (G, T, α)
- $S := \{1..Smax\}$: set of stages and/or mutants.
- $U = \{U_s\}_{s \in S}; U_s \subseteq V$: nodes of G_s, the (induced) subnetworks of G.
- $R = \{R_s\}_{s \in S}; R_s := \{1..Rmax_s\}$: regions of homogeneous expression.
- $I = \{\imath_{s,r,u}\}_{s \in S, r \in R_s, u \in U_s}; \imath_{s,r,u} : V \rightarrow \{0,1\}$: initial conditions.
- $\Phi = \{\phi_{s,r,u}\}_{s \in S, r \in R_s, u \in U_s}; \phi_{s,r,u} : V \rightarrow \{0,1\}$: expression data.

Find

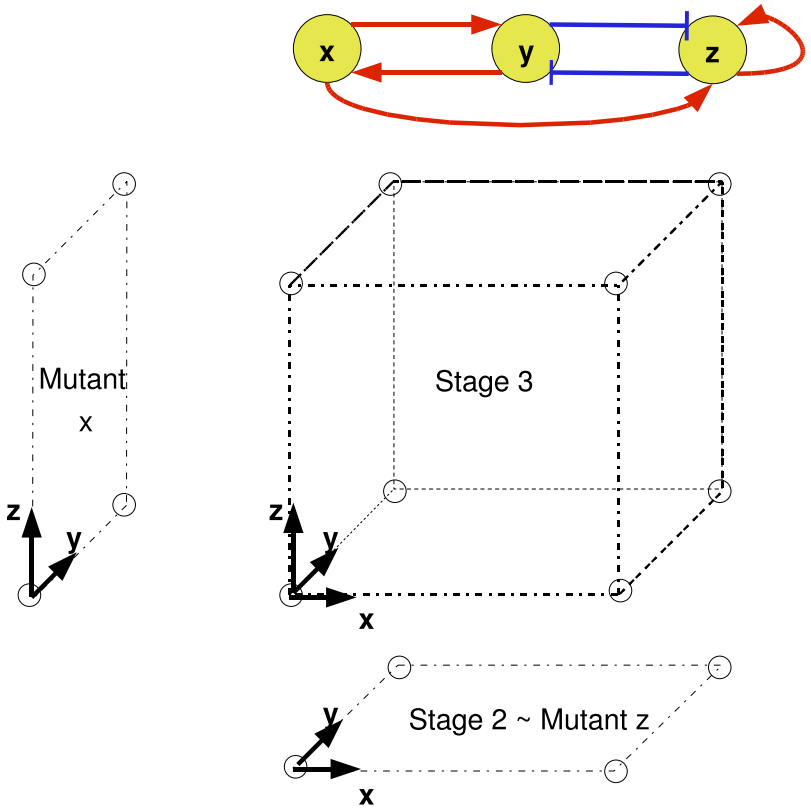
w, θ such that $\forall\ \vec{\imath}_{s,r}$, $(G_s, T, \alpha, w, \vec{x}_{s,r}, \vec{\imath}_{s,r}, \theta)$ satisfies the evolution constraints and have fixed points that collectively minimize $D_H(\vec{\rho}, \vec{\phi})$.

$D_H(\vec{\rho}, \vec{\phi})$: total hamming distance from model fixed points to data.

fixed points ($\vec{x}$) : If $\vec{x}$ $\vec{x}$ $\vec{x}$ then $\vec{x}$ $\vec{x}$ for all $t' > t$

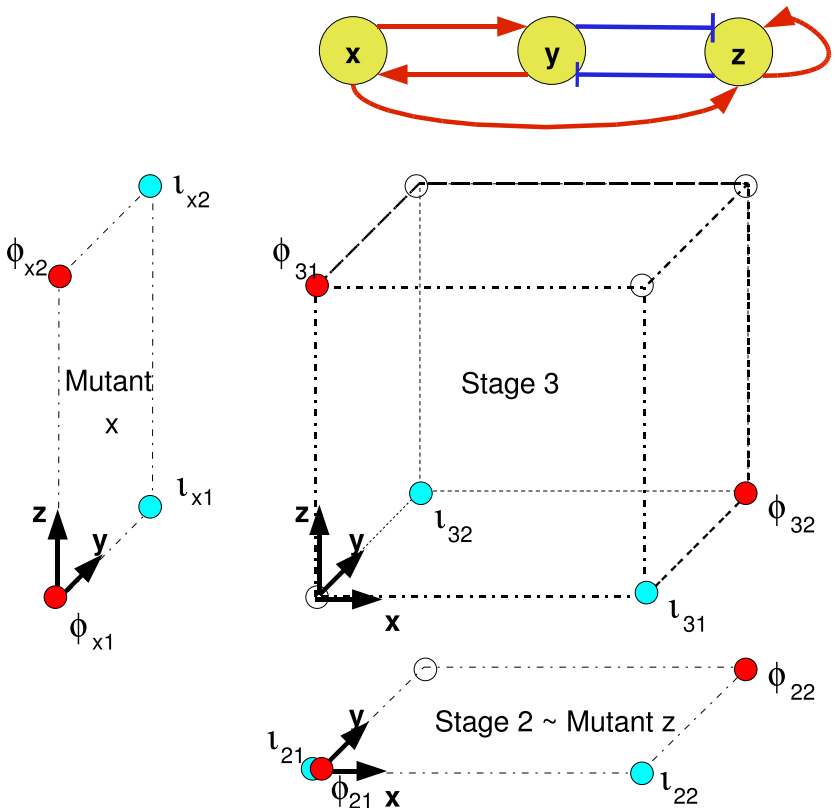
Modelling the inverse problem: ilustrating the problem

Finding network parameters for simulaneous stable subnetworks



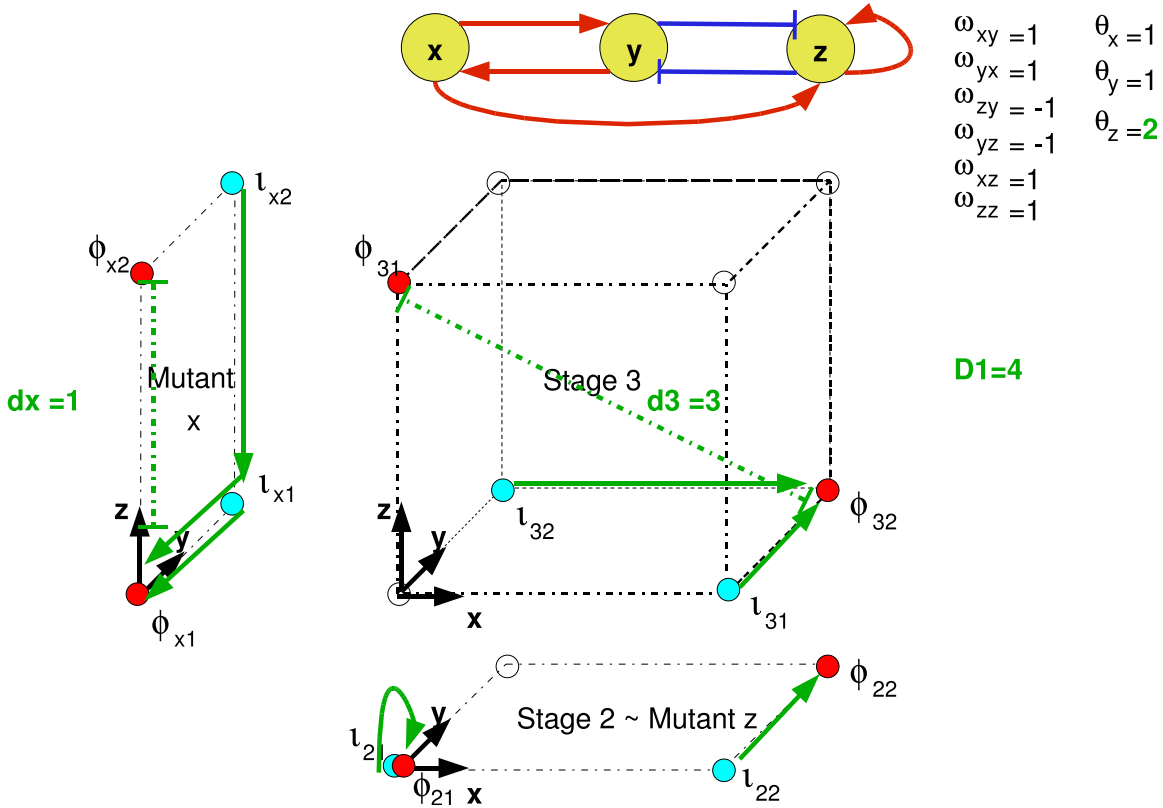
Modelling the inverse problem: ilustrating the problem

Finding network parameters for simulaneous stable subnetworks



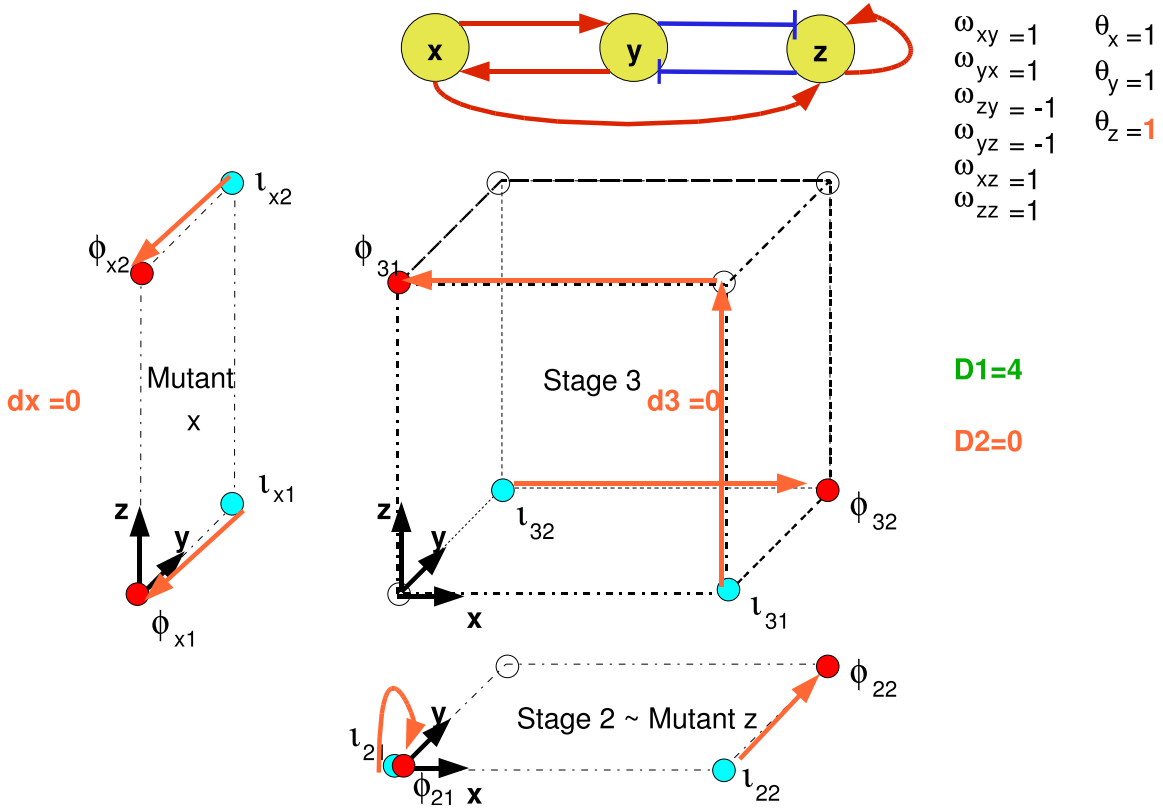
Modelling the inverse problem: ilustrating the problem

Finding network parameters for simulaneous stable subnetworks



Modelling the inverse problem: ilustrating the problem

Finding network parameters for simulaneous stable subnetworks



Mathematical Programming Formulation

$$\left.\begin{array}{rrcl} \min_x & f(x) & & \\ \text{subject to} & g(x) & \leq & 0, \end{array}\right\}$$

x: decision variables, f: objective function, g : constraints

Sets	V, A, T, S, R (genes, interactions, time steps, stages, regions)
Variables	$x : V \times R \times T \rightarrow \{0,1\}$, $w : A \rightarrow \mathbb{R}^+$, $\theta : A \rightarrow \mathbb{R}$
Parameters	$\alpha : A \rightarrow \{-1,+1\}$, bounds: $\theta^L, \theta^U, w^L, w^U$ $\phi_{v,r}$, $\iota_{v,r}$ (observed gene expression and initial cond.)

Objective Function and Constraints

Objective function

$$\sum_{s\in S, r\in R_s} \sum_{t\in T\setminus 1} (\sigma_{s,r}^{t-1} - \sigma_{s,r}^{t}) \sum_{u\in U_s} \left| x_{s,u,r}^{t} - \rho_{s,u,r} \right|$$

Evolution rules

$$\sum_{u\in U_s:(u,v)\in A} \alpha_{u,v} w_{u,v} x_{s,r,u}^{t-1} \geq \theta_v x_{s,r,v}^{t} - \|V\| (1 - x_{s,r,v}^{t})$$

$$\sum_{u\in U_s:(u,v)\in A} \alpha_{u,v} w_{u,v} x_{s,r,u}^{t-1} \leq (\theta_v - \varepsilon)(1 - x_{s,r,v}^{t}) + \|V\| x_{s,r,v}^{t}$$

Fixed point conditions

$$\sum_{u\in U_s} |x_{s,u,r}^{t} - x_{s,u,r}^{t-1}| \leq \|U_s\| \sigma_{s,r}^{t}$$

$$\sum |x_{s,u,r}^{t} - x_{s,u,r}^{t-1}| \geq \sigma_{s,r}^{t}$$

Reformulation, Linearization and Solution

Nonconvex Mixed-Integer Nonlinear Program (MINLP).
Reformulated exactly to a MILP.

yx terms	θx terms
(y, x: binary)	(θ: real, x: binary)
$z \geq 0$	$\zeta \geq \theta^L x$
$z \leq y$	$\zeta \leq \theta + (\lvert\theta^L\rvert + \lvert\theta^U\rvert)(1-x)$
$z \leq x$	$\zeta \leq \theta^U x$
$z \geq x + y - 1$	$\zeta \geq \theta - (\lvert\theta^L\rvert + \lvert\theta^U\rvert)(1-x)$

- Absolute values and distances.
- Auxiliary decision variables for fixed point conditions.

We use AMPL to write the model of the problem, and
use CPLEX 11.0.1 to solve efficiently to optimality the MILP problem.

Ongoing work: transitory dynamics

- Deterministic vs Stochastic
- Deterministic: Asynchronouse vs synchronous
- Biological interpretation

Asynchronous parameters

$$x_i^{(p_i\tau+q_i+1)} = F_i(x^{(p_i\tau+q_i)})$$

- p_i: period
- q_i: delay

Biologically based

$$\tau_{x_i^0} = \lambda_i log(\frac{1}{1-d_i(x_i^0)/D_i(x_i^0))})$$

- λ_i : gene product degradation
- D_i : distance to state image
- d_i : distance to threshold

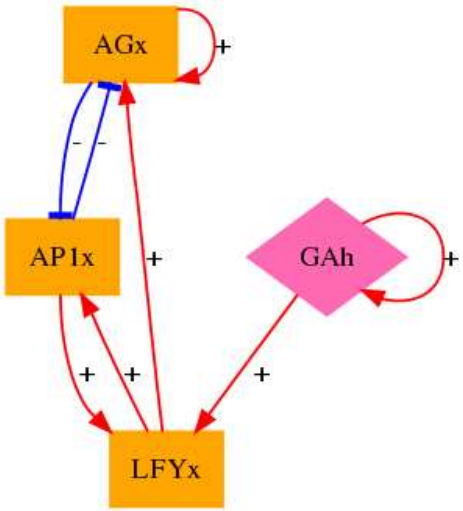

AGx
AP1x
GAh
LFYx
+
+
+
+
+
+
-
-

Figure: WUS mutant

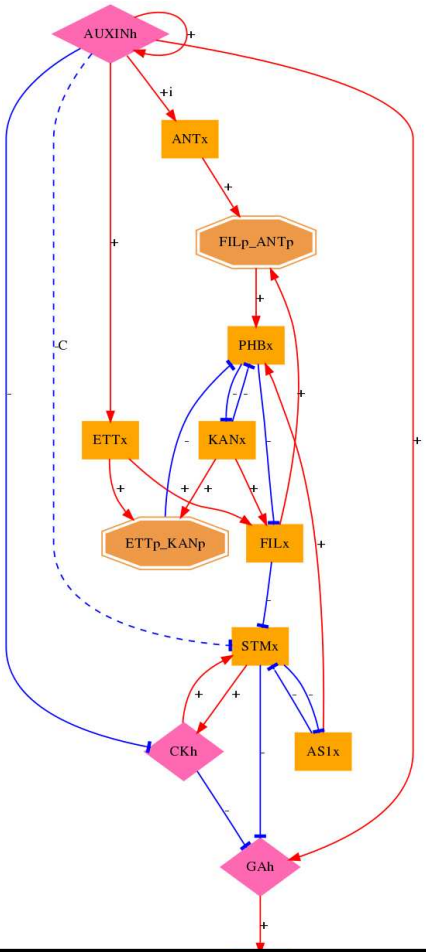


Figure: Wild-type Stage 2

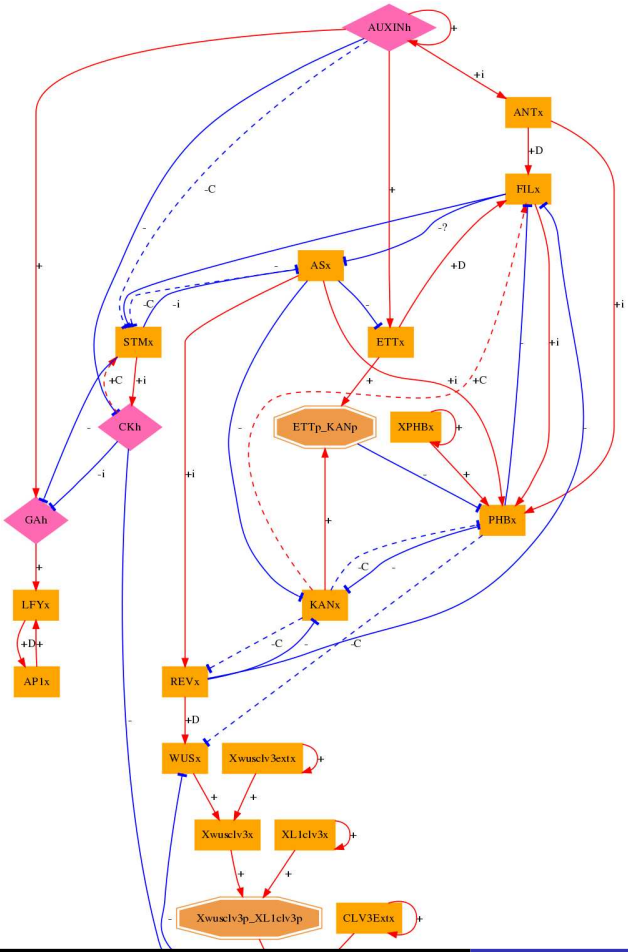
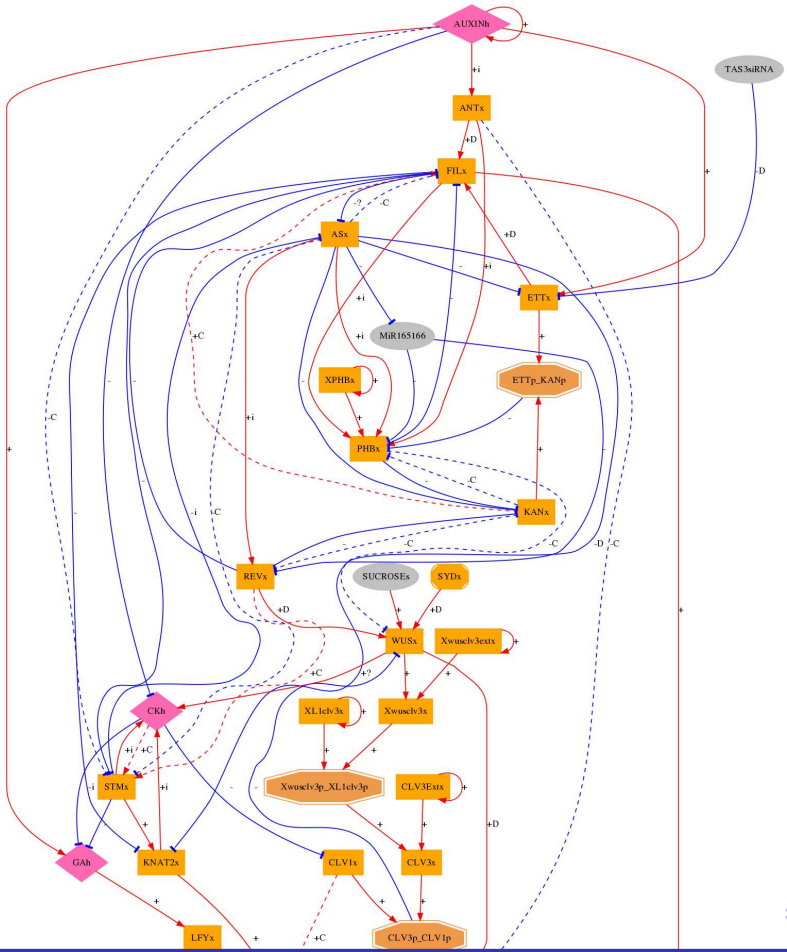


Figure: Wild-type Stage 3



Summary

- Static modelling of a dynamic system.
- Generic modeling approach for the inference of biological regulatory networks.
- Easier to test different models than simulation approaches.

- Perspectives
  - “Flexibilize” the “hard” constraint on the prior network (find signs, new interactions)
  - Introduce theoretical results on regulatory networks.
  - Multiobjective problems ?
  - Reintroduce transitory dynamics (Is it possible using mathematical programming ?).
  - Study more complicated qualitative models of GRN.