Introduction to Bioinformatics

Systems biology: modeling biological networks

Systems biology

- Study of ”whole biological systems”
- ”Wholeness”: Organization of dynamic interactions
  - Different behaviour of the individual parts when isolated or when combined together
  - Systems cannot be fully understood by analysis of their components in isolation

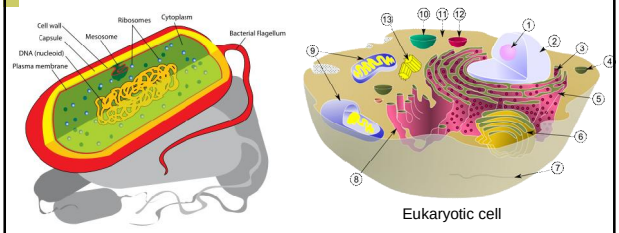
-- Ludwig von Bertalanffy, 1934
(according to Zvelebil & Baum)

Outline

- 1. Systems biology and biological networks
  - Transcriptional regulation
  - Metabolism
  - Signalling networks
  - Protein interactions
- 2. Modeling frameworks
  - Continuous and discrete models
  - Static and dynamic models
- 3. Identification of models from data

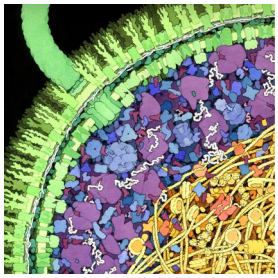
1. Systems biology

- Systems biology – biology of networks
  - Shift from component-centered biology to systems of interacting components



http://en.wikipedia.org/wiki/Cell_(biology)
Mariana Ruiz, Magnus Manske

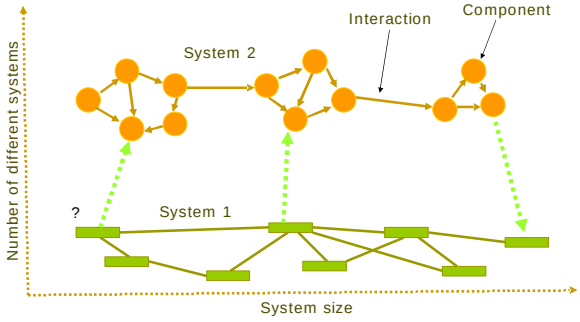
Interactions within the cell



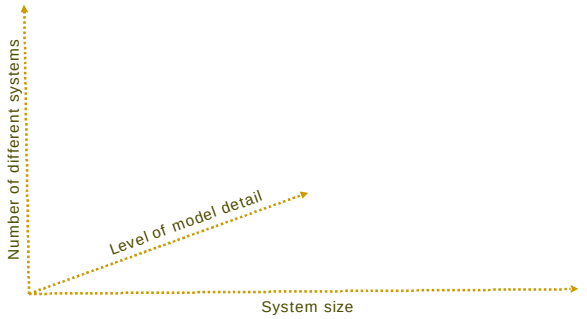
http://mgl.scripps.edu/people/goodsell/illustration/public
David S. Goodsell

- Density of biomolecules in the cell is high: plenty of interactions!
- Figure shows a cross-section of an Escherichia coli cell
  - Green: cell wall
  - Blue, purple: cytoplasmic area
  - Yellow: nucleoid region
  - White: mRNA

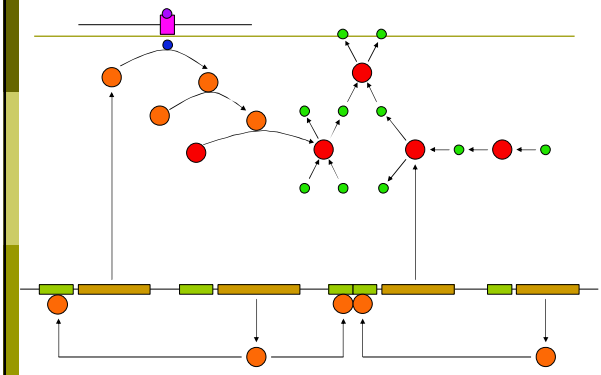
Paradigm shift from study of individual components to systems



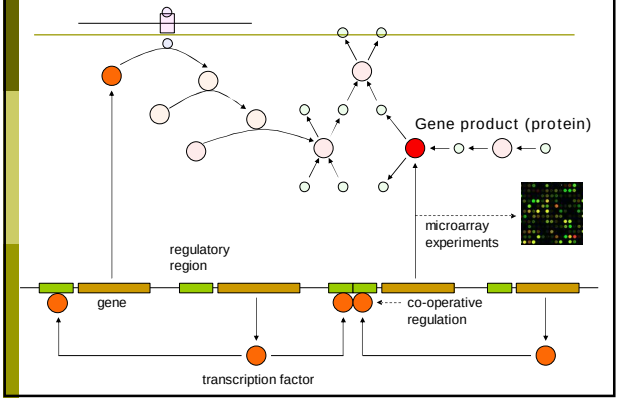
Paradigm shift from study of individual components to systems



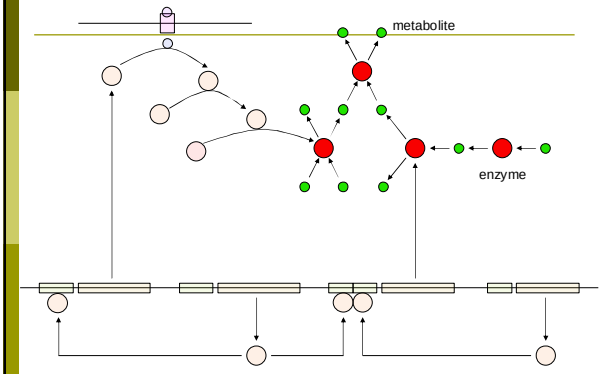
Biological systems of networks



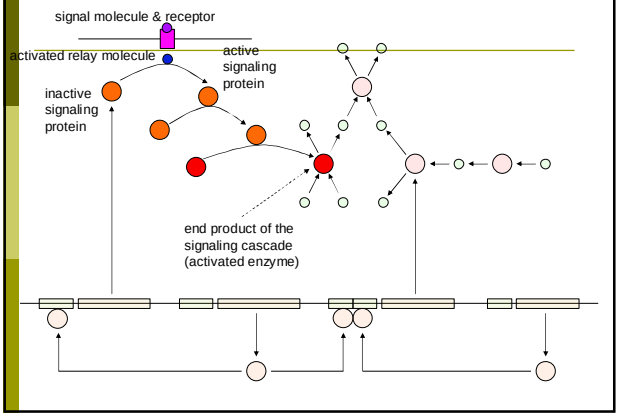
Transcriptional regulation



Metabolism



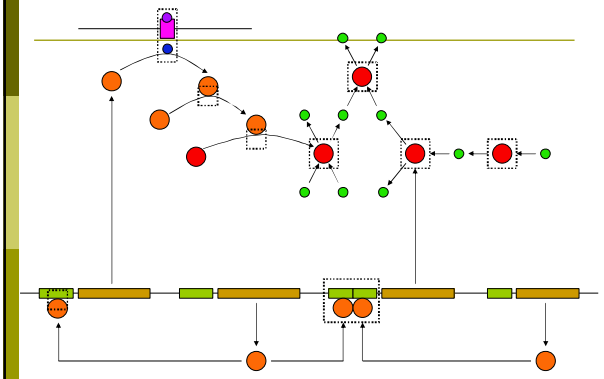
Signal transduction



Protein interaction networks

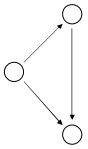
- Protein interaction is the unifying theme of all regulation at the cellular level
- Protein interaction occurs in every cellular system including systems introduced earlier
- Data on protein interaction reveals associations both within a system and between systems

Protein interaction



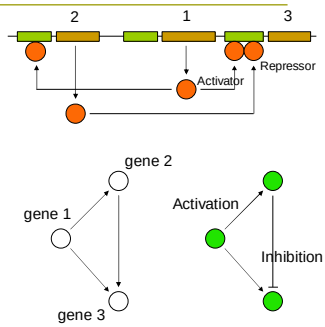
2. Graphs as models of biological networks

- A graph is a natural model for biological systems of networks
- Nodes of a graph represent biomolecules, edges interactions between the molecules
- Graph can be undirected or directed
- To address questions beyond simple connectivity (node degree, paths), one can *enrich* the graph models with information relevant to the modeling task at hand

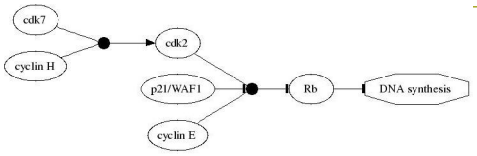


Enriching examples: transcriptional regulation

- Regulatory effects can be (roughly) divided into
  - activation
  - inhibition
- We can encode this distinction by labeling the edges by '+' and '-', for example
- Graph models of transcriptional regulation are called *gene(tic) regulatory networks*



Enriching examples: more transcriptional regulation



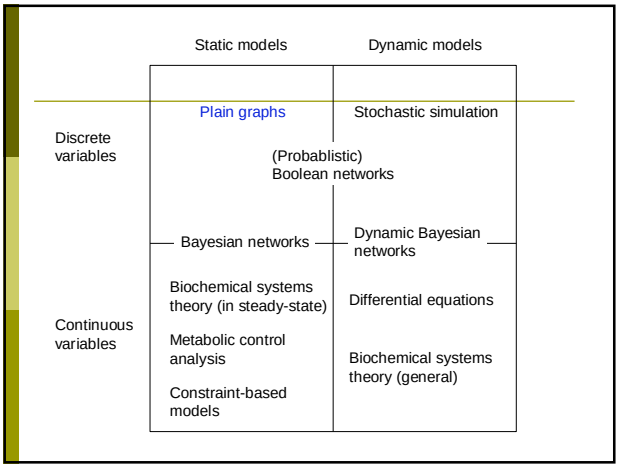
A gene regulatory network might be enriched further: In this diagram, proteins working cooperatively as regulators are marked with a black circle.

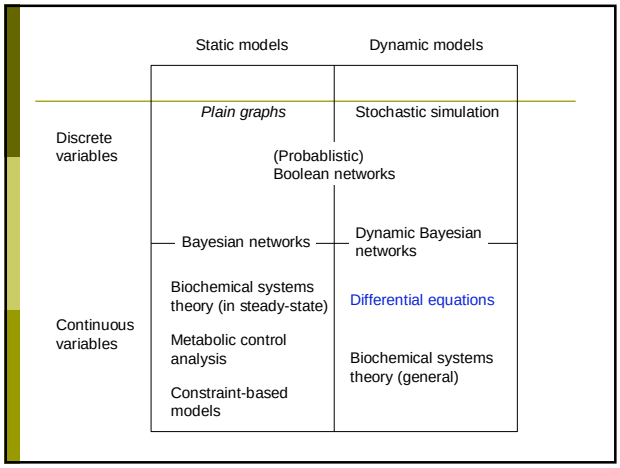
This network is a simplified part of cell cycle regulation.

Frameworks for biological network modeling

- A variety of information can be encoded in graphs
- Modeling frameworks can be categorised based on what sort of information they include
  - Continuous and/or discrete variables?
  - Static or dynamic model? (take time into account?)
  - Spatial features? (consider the physical location molecules in the cell?)
- Choice of framework depends on what we want to do with the model:
  - *Data exploration*
  - *Explanation of observed behaviour*
  - *Prediction*

	Static models	Dynamic models
Discrete variables		
Continuous variables		





Dynamic models: differential equations

- In a differential equation model
  - variables x_i correspond to the concentrations of biological molecules;
  - change of variables over time is governed by rate equations,

$$dx_i/d_t = f_i(x),\ 1 \le i \le n$$

- In general, $f_i(x)$ is an arbitrary function (not necessarily linear)
- Note that the graph structure is encoded by parameters to functions $f_i(x)$

Properties of a differential equation model

- The crucial step in specifying the model is to choose functions $f_i(x)$ to balance
  - model complexity (number of parameters)
  - level of detail
- Overly complex model may need more data than is available to specify

Example of a differential equation model of transcriptional regulation

- Let x be the concentration of the target gene product
- A simple kinetic (i.e., derived from reaction mechanics) model could take into account
  - multiple regulators of target gene and
  - degradation of gene products

  and assume that regulation effects are independent of each other

Example of a differential equation model of transcriptional regulation

- Rate equation for change of x could then be

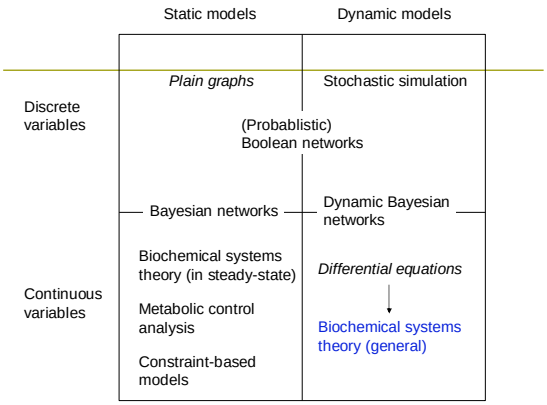
$$\frac{dx}{dt} = k_1 \frac{1}{1 + \exp(-\sum_{j=1..m} w_j y_j + b)} - k_2 x$$

  where k_1 is the maximal rate of transcription of the gene, k_2 is the rate constant of target gene degradation, w_j is the regulatory weight of regulator j and y_j is the concentration of regulator j

Number of parameters?

Differential equation model for metabolism

- Likewise, rate equations can be derived for differential equation models for metabolism
- For simple enzymes, two parameters might be enough
- Realistic modeling of some enzyme requires knowledge of 10-20 parameters
- Such data is usually not available in high-throughput manner

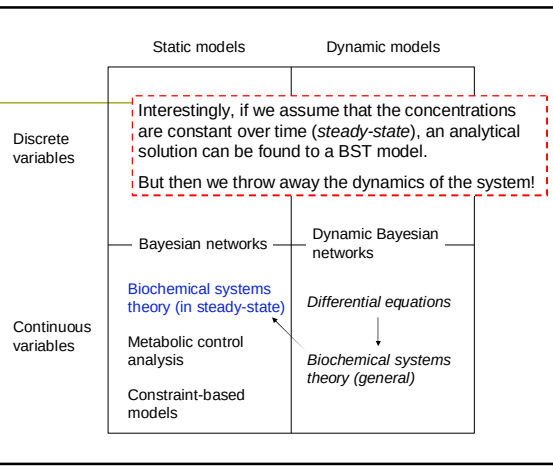


Biochemical systems theory (BST)

- BST is a modeling framework, where differential rate equations are restricted to the following power-law form,

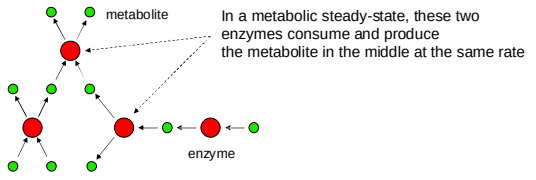
$$\frac{dx_i}{dt} = \alpha_i \prod_{j=1}^{n} x_j^{g_{ij}}$$

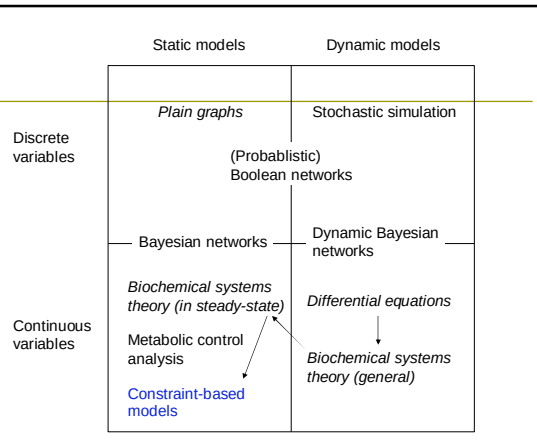
where
  - α_i is the rate constant for molecule i and
  - g_{ij} is a kinetic constant for molecule i and reaction j
- BST approximates the kinetic system and requires less parameters than the genetic kinetic model



Steady-state modeling

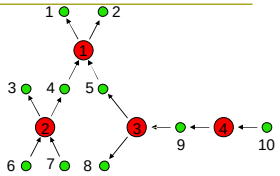
- Is the study of steady-states meaningful?
- If we assume $dx_i/dt = 0$, we restrict ourselves to systems, where the production of a molecule is balanced by its consumption





Constraint-based modeling

- Constraint-based modeling is a linear framework, where the system is assumed to be in a steady-state
- Model is represented by a *stoichiometric matrix* S, where S_{ij} gives the number of molecules of type i produced in reaction j in a time unit.



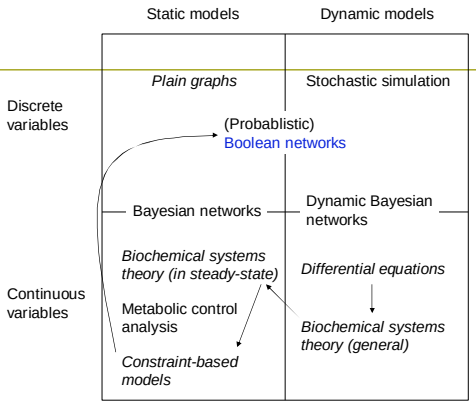
	1	2	3	4
1	1			
2	1			
3		2		
4	-1	1		
5	-1		1	
6		-2		
7		-1		
8			1	
9			-2	1
10				-1

S_{ij} = 0 if value omitted

Constraint-based modeling

- Since variables x_i are constant, the questions asked now deal with reaction rates
- For instance, we could characterise solutions to the linear steady-state condition, which can be written in matrix notation as

  $Sv = 0$
- Solutions v are reaction rate vectors, which for example reveal alternative pathways inside the network



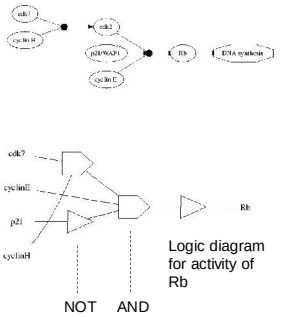
Discrete models: Boolean networks

- Boolean networks have been widely used in modeling gene regulation
  - Switch-like behaviour of gene regulation resembles logic circuit behaviour
  - Conceptually easy framework: models easy to interpret
  - Boolean networks extend naturally to dynamic modeling

Boolean networks

A Boolean network G(V, F) contains

- Nodes $V = \{x_1, \ldots, x_n\}$, $x_i = 0$ or $x_i = 1$
- Boolean functions $F = \{f_1, \ldots, f_n\}$
- Boolean function f_i is assigned to node x_i



Logic diagram for activity of Rb

Dynamics in Boolean networks

- Dynamic behaviour can be simulated
- State of a variable x_i at time t+1 is calculated by function f_i with input variables at time t
- Dynamics are deterministic: state of the network at any time depends only on the state at time 0.

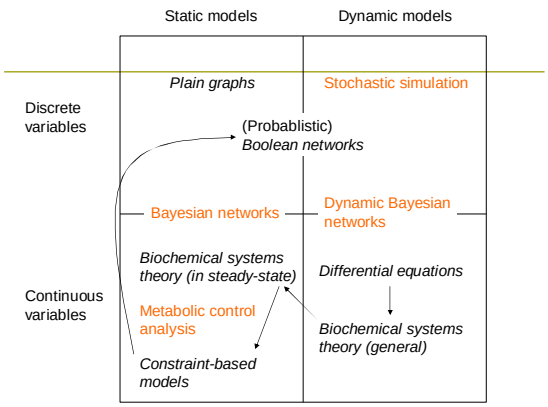
Example of Boolean network dynamics

- Consider a Boolean network with 3 variables x_1, x_2 and x_3 and functions given by
  - $x_1 := x_2$ and x_3
  - $x_2 :=$ not x_3
  - $x_3 := x_1$ or x_2

t	x1	x2	x3
0	0	0	0
1	0	1	0
2	0	1	1
3	1	0	1
4	0	0	1
		...	

Problems with Boolean networks

- 0/1 modeling is unrealistic in many cases
- Deterministic Boolean network does not cope well with missing or noisy data
- Many Boolean networks to choose from – specifying the model requires a lot of data
  - A Boolean function has n parameters, or inputs
  - Each input is 0 or 1 => 2^n possible input states
  - The function is specified by input states for which f(x) = 1 => 2^(2^n) possible Boolean functions



3. Model identification from data

- We would like to learn a model from the data such that the learned model
  - Explains the observed data
  - Predicts the future data well
- Generalization property: model has a good tradeoff between a good fit to the data and model simplicity

Three steps in learning a model

- *Representation*: choice of modeling framework, how to encode the data into the model
  - Restricting models: number of inputs to a Boolean function, for example
- *Optimization*: choosing the "best" model from the framework
  - Structure, parameters
- *Validation*: how can one trust the inferred model?

Conclusions

- Graph models are important tools in systems biology
- Choice of modeling framework depends on the properties of the system under study
- Particular care should be paid to dealing with missing and incomplete data - choice of the framework should take the quality of data into account

References and further reading

- Florence d'Alché-Buc and Vincent Schachter: Modeling and identification of biological networks. In Proc. Intl. Symposium on Applied Stochastic Models and Data Analysis, 2005.
- Marketa Zvelebil and Jeremy O. Baum: Understanding bioinformatics. Garland Science, 2008.
- Hiroaki Kitano: Systems Biology: A Brief Overview. Science 295, 2002.
- Marie E. Csete and John C. Doyle: Reverse engineering of biological complexity. Science 295, 2002.
- James M. Bower and Hamid Bolouri (eds): Computational Modeling of Genetic and Biochemical Networks. MIT Press, 2001.