Modeling Optimal Gene Regulatory Networks

supervisor: Professor Jerzy Tiuryn

Andrzej Mizera

am189425@students.mimuw.edu.pl

Faculty of Mathematics, Informatics, and Mechanics
Warsaw University

Institute of Fundamental Technological Research

Polish Academy of Sciences

Outline

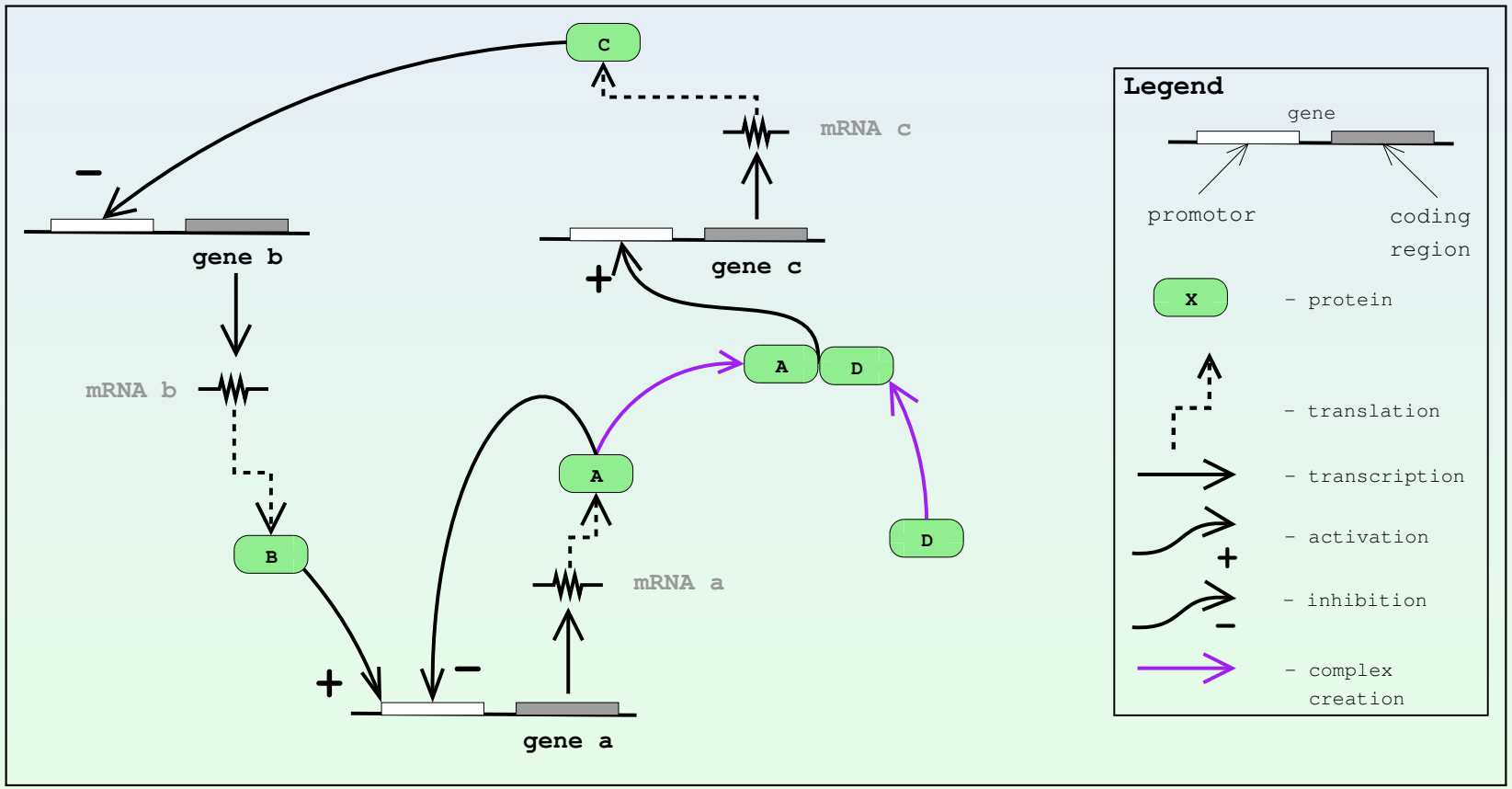
Outline

Outline

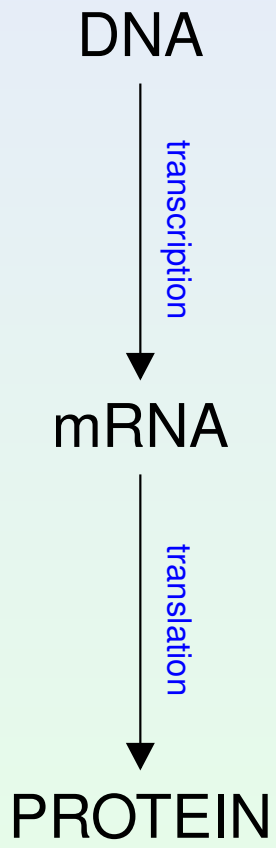
Outline

Definition

gene regulatory network - a collection of DNA segments in a cell which interact with each other and with other substances in the cell, thereby governing the rates at which genes in the network are transcribed into mRNA.



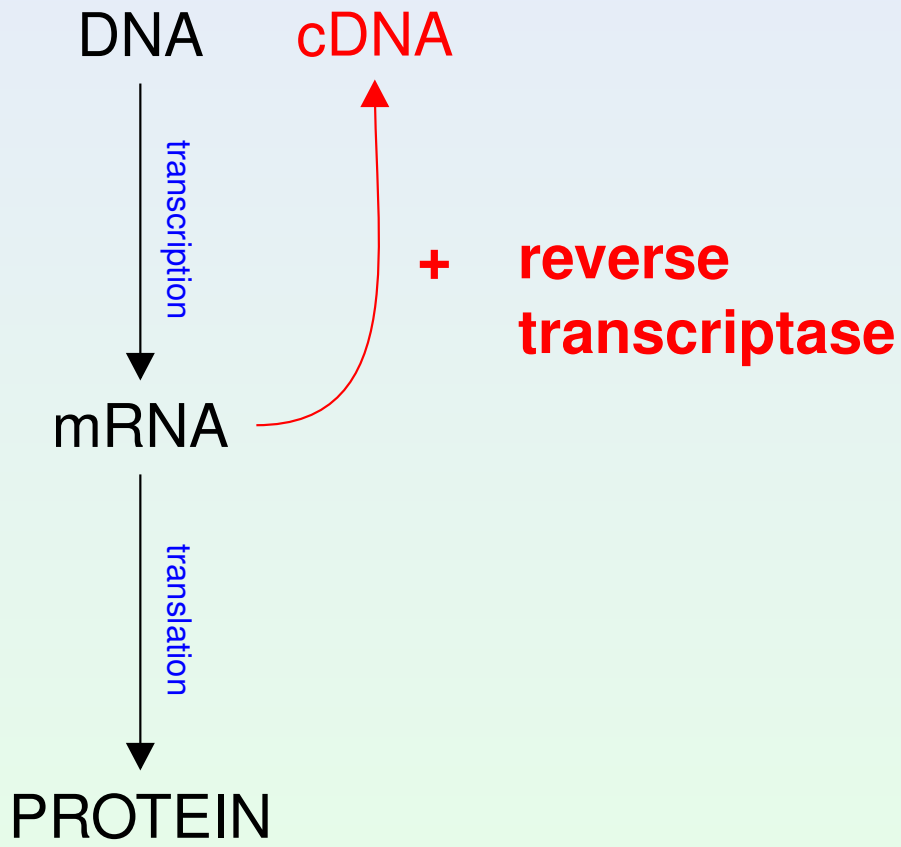
The Central Dogma of Molecular Biology



Characteristics

- The central dogma of molecular biology deals with the transfer of sequential information.
- It states that such information cannot be transferred from protein to either protein or nucleic acid.
- Three groups of transfers:
  - **general transfers** (believed to occur normally in most cells),
  - **special transfers** (known to occur, but only under abnormal conditions),
  - **unknown transfers** (believed never to occur).
- The general transfers describe the normal flow of biological information: DNA information can be copied into mRNA and proteins can be synthesized using the information in mRNA as a template.

The Central Dogma of Molecular Biology



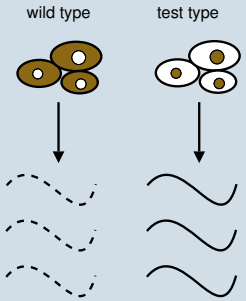
Characteristics

- The central dogma of molecular biology deals with the transfer of sequential information.
- It states that such information cannot be transferred from protein to either protein or nucleic acid.
- Three groups of transfers:
  - **general transfers** (believed to occur normally in most cells),
  - **special transfers** (known to occur, but only under abnormal conditions),
  - **unknown transfers** (believed never to occur).
- The general transfers describe the normal flow of biological information: DNA information can be copied into mRNA and proteins can be synthesized using the information in mRNA as a template.

DNA Microarray

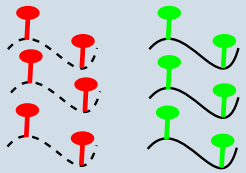
Description

A) Isolation of mRNA (the cells have grown and ascertained which genes had to be activated or repressed in order fot the cell to survive).



B) Synthesis of cDNA from mRNA with reverse transcriptase.

C) Labeling cDNA by fluorescent dye (wild type - red, test type - green).



D) DNA microarray (DNA chip) consists of spots. Each spot is made of gene specific DNA that can base pair with cDNA fragments.

E) cDNA hybridization to DNA Microarray spots.

F) Scanning with a green and then a red laser in order to detect the bounded cDNA.

G) Image marging (computer analysis).

- wild type concentration $>$ test type concentration (repression)
- wild type concentration $=$ test type concentration
- wild type concentration $<$ test type concentration (activation)

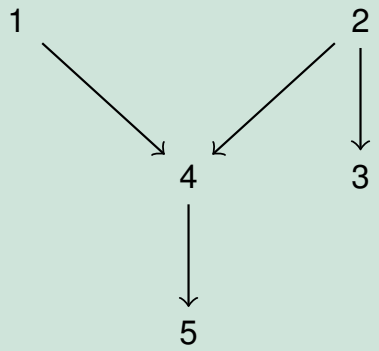
Definition

Bayesian network $\mathcal{B} = \langle G, \Theta \rangle$ is a representation of a joint probability distribution over a set of random variables $\mathcal{X}$. It consists of two components:

- G – a directed acyclic graph whose vertices correspond to random variables and edges indicate conditional dependence relations,
- Θ – a family of conditional distributions for each variable, given its parents in the graph.

Example

Bayesian network (structure + conditional probability table (CPT))



Θ_{X_4}
$Pr(X_4 = 0 \mid X_1 = 0, X_2 = 0) = 0.30$ $Pr(X_4 = 1 \mid X_1 = 0, X_2 = 0) = 0.70$
$Pr(X_4 = 0 \mid X_1 = 0, X_2 = 1) = 0.76$ $Pr(X_4 = 1 \mid X_1 = 0, X_2 = 1) = 0.24$
$Pr(X_4 = 0 \mid X_1 = 1, X_2 = 0) = 0.12$ $Pr(X_4 = 1 \mid X_1 = 1, X_2 = 0) = 0.88$
$Pr(X_4 = 0 \mid X_1 = 1, X_2 = 1) = 0.95$ $Pr(X_4 = 1 \mid X_1 = 1, X_2 = 1) = 0.05$

Joint probability distribution

The graph structure G encodes the following set of independence assumptions: each node X_i is independent of its non-descendants given its parents in G.

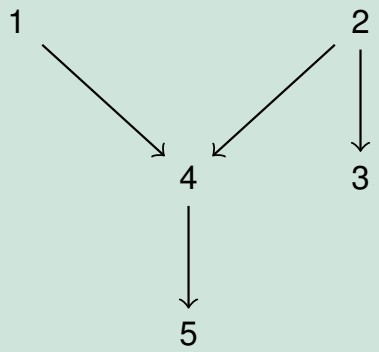
The joint probability distribution can be expressed in the following way according to the Chain Rule (independent of the ordering) and encoded set of independencies:

$$P(\mathcal{X}) = P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | X_1, \ldots, X_{i-1}) = \prod_{i=1}^{n} P(X_i | Pa(X_i)).$$

Bayesian network $\mathcal{B}$ defines a unique joint probability distribution over $\mathcal{X}$.

Example

Independencies encoded by the structure of a Bayesian network



$i(X_1; X_2, X_3)$,
$i(X_2; X_1)$,
$i(X_3; X_1, X_4, X_5 | X_2)$,
$i(X_4; X_3 | X_1, X_2)$,
$i(X_5; X_1, X_2, X_3 | X_4)$

Equivalence Classes of Bayesian Networks

Definition

Two graphs G and G' with the same set of nodes ($V = V'$) are equivalent if for each Bayesian network $\mathcal{B} = \langle G, \Theta \rangle$ there exist another Bayesian network $\mathcal{B}' = \langle G', \Theta' \rangle$ such that both $\mathcal{B}$ and $\mathcal{B}'$ define the same joint probability distribution and vice versa.

Theorem (Pearl, and Verma, 1991)

Two graphs are equivalent if and only if their DAGs have the same underlying undirected graph and the same v-structures (converging arrows emanating from non-adjacent nodes).

Caution

On the basis of observations from a distribution one cannot distinguish between equivalent graphs!

Equivalence Classes of Bayesian Networks

Definition

Two graphs G and G' with the same set of nodes ($V = V'$) are equivalent if for each Bayesian network $\mathcal{B} = \langle G, \Theta \rangle$ there exist another Bayesian network $\mathcal{B}' = \langle G', \Theta' \rangle$ such that both $\mathcal{B}$ and $\mathcal{B}'$ define the same joint probability distribution and vice versa.

Theorem (Pearl, and Verma, 1991)

Two graphs are equivalent if and only if their DAGs have the same underlying undirected graph and the same v-structures (converging arrows emanating from non-adjacent nodes).

Caution

On the basis of observations from a distribution one cannot distinguish between equivalent graphs!

Equivalence Classes of Bayesian Networks

Definition

Two graphs G and G' with the same set of nodes ($V = V'$) are equivalent if for each Bayesian network $\mathcal{B} = \langle G, \Theta \rangle$ there exist another Bayesian network $\mathcal{B}' = \langle G', \Theta' \rangle$ such that both $\mathcal{B}$ and $\mathcal{B}'$ define the same joint probability distribution and vice versa.

Theorem (Pearl, and Verma, 1991)

Two graphs are equivalent if and only if their DAGs have the same underlying undirected graph and the same v-structures (converging arrows emanating from non-adjacent nodes).

Caution

On the basis of observations from a distribution one cannot distinguish between equivalent graphs!

Drawbacks of standard Bayesian Networks

- Existence of equivalence classes of Bayesian networks creates problems in assigning direction of causation to an interaction.
- Due to the mathematical properties of the joint probability distribution Bayesian networks have to be acyclic. This restriction causes problems in applications of this formalism in biology, because feedback loops are a common biological feature.

Both of these limitations can be overcome by using **Dynamic Bayesian Networks**.

The formalism of Dynamic Bayesian Networks (DBNs)

Description

- Dynamic Bayesian Networks are directed graphical models of stochastic processes.
- Represented process is assumed to satisfy the Markovian condition, i.e.

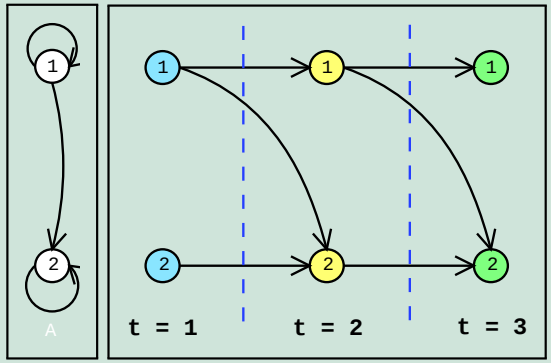
$$P(\mathcal{X}(t)|\mathcal{X}(0), \mathcal{X}(1), \ldots, \mathcal{X}(t-1)) = P(\mathcal{X}(t)|\mathcal{X}(t-1))$$

- and to be time homogenous, i.e.

$$P(\mathcal{X}(t)|\mathcal{X}(t-1))$$

  are independent of t.
- A Dynamic Bayesian Network consists of a graph G and a family of parameters Θ which characterise the conditional probability distributions $P(X_i(t)|Pa(X_i)(t-1))$, where $X_i \in \mathcal{X}$.

Example

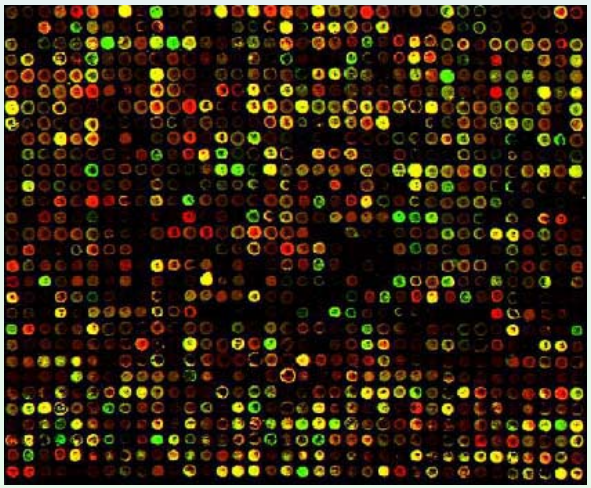


An example of a Dynamic Bayesian Network (left figure) and the same network unwrapped in time (right figure). The unwrapped network is acyclic.

Statement of the problem

Problem Statement

Given a training set $D = \{\boldsymbol{x_1}, \ldots, \boldsymbol{x_N}\}$ of independent instances of $\mathcal{X}$, find a network $\mathcal{B}' = \langle G, \Theta \rangle$ that best matches D (more precisely, the equivalence class of networks that best matches D).



Experiments:	***1***	***2***	***3***	***4***	***5***	***6***
X_1	-1	1	0	1	0	0
X_2	0	-1	0	1	0	-1
X_3	0	1	-1	1	0	-1

Solution

- Introduce a statistically motivated scoring function that evaluates each network with respect to the training data.
  A commonly used scoring is the Bayesian score:

$$Score(G, D) = \log P(G|D) = \log P(D|G) + \log P(G) + C,$$

  where C is a constant independent of G and

$$P(D|G) = \int P(D|G, \Theta)P(\Theta|G)d\Theta.$$

- Learning amounts to finding the structure G that maximizes or minimizes the score.
- The Bayesian score class of functions realises the Maximum a posteriori rule - the graph G that maximizes $P(G|D)$ is chosen.
- This problem is NP-hard - usually heuristic methods are used.
- The decomposition ot the score is crucial for this optimization problem.

Definition

A score function is decomposable if it can be rewritten as the sum

$$Score(G, D) = \sum_i ScoreContribution(X_i, Pa(X_i), D),$$

where the contribution of every variable X_i to the total network score depends only on its own value and the values of its parents in G.

The algorithm

S. Ott, S. Imoto, and S. Miyano
Finding Optimal Models for Small Gene Networks.
Pacific Symposium on Biocomputing, 9:557-567, 2004.

Algorithm – the idea

part I: for each gene $g \in G$

for each potential parent set *Pa* of *g*

compute the local score for *g* and *Pa*

part II: on the basis of previous computations choose the parent set for each *g* yielding the optimal score of the whole network

Characteristics of the algorithm

- Decomposition of the score function is crucial for this algorithm.
- Exhaustive search is performed – elimination of heuristics.
- Time complexity:
  - part I: $O(n2^n)$ operations of computing $score(gene, parents)$
  - part II: $O(n2^n)$ operations are needed.
- The dynamic programming approach is used.

- Both standard and dynamic formalisms of Bayesian Networks.
- Two scoring functions:

  MDL (Minimal Description Length) – has a simple motivation in universal coding.

  The description length of the data based on a model
  $=$
  length of the compressed data
  $+$
  the representation size of the model itself.

  MDL principle dictates that the optimal model is the one that minimizes the total description length.

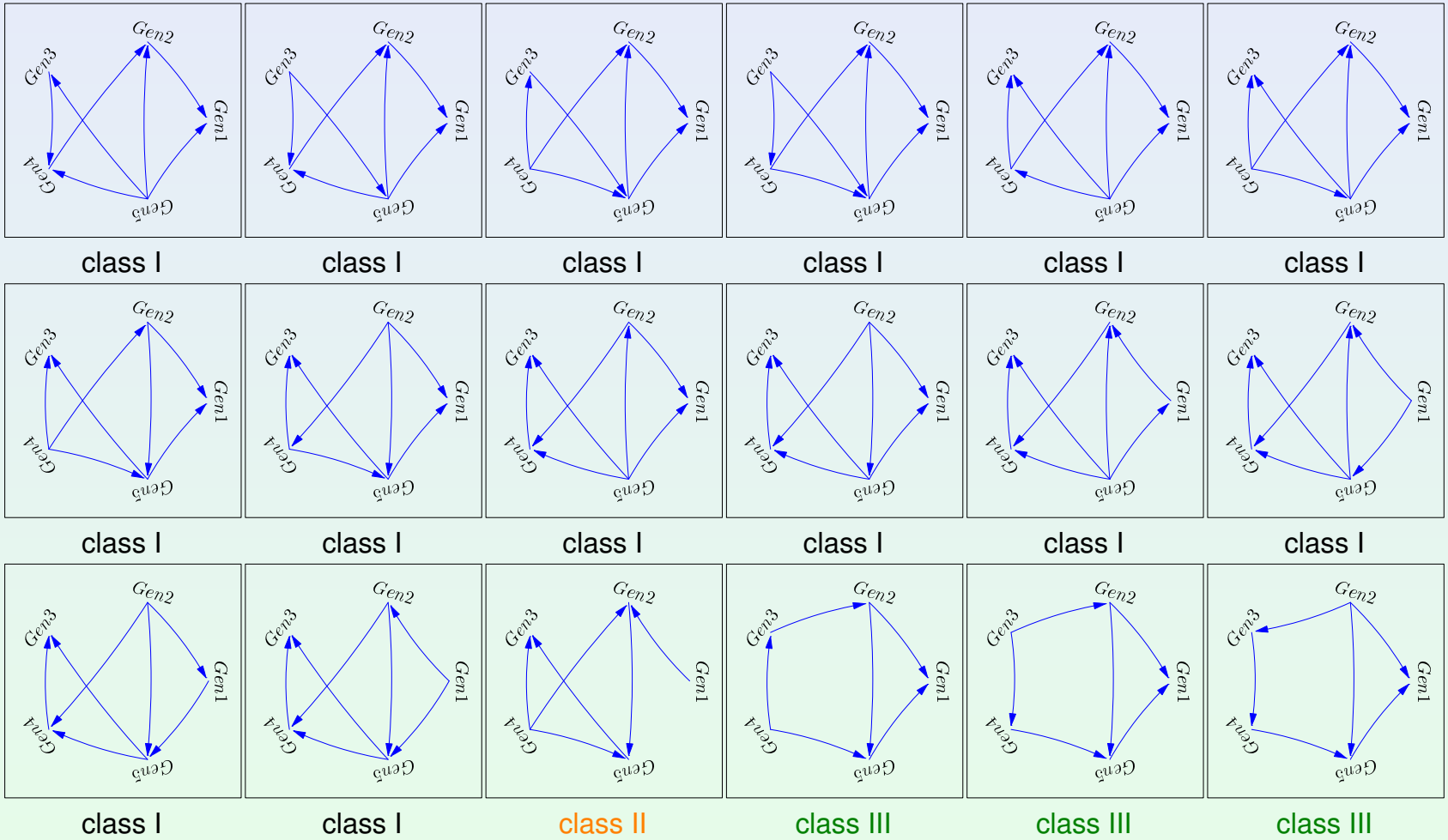
  BDe (Bayesian Dirichlet equivalence) – derived from the posterior probability of the network, given the data.

- Finding the structure of all optimal network structures – the class of optimal networks.
- Finding the structure of all networks in a requested number of suboptimal classes.
- Bulding a consensus network (containing n most conserved egdes) from computed networks.
- Pearson's correlation coefficient is calculated between X_i and X_j iff $(v_i, v_j) \in E$.

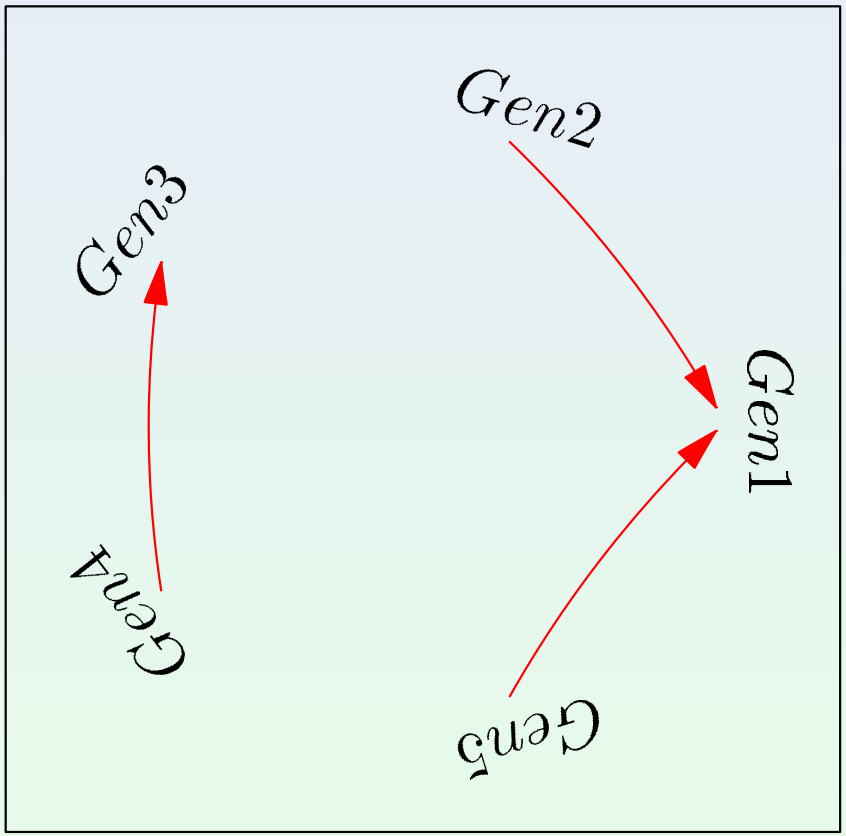
  $\sigma_{X_i X_j} < 0$ inhibition

  $\sigma_{X_i X_j} > 0$ activation

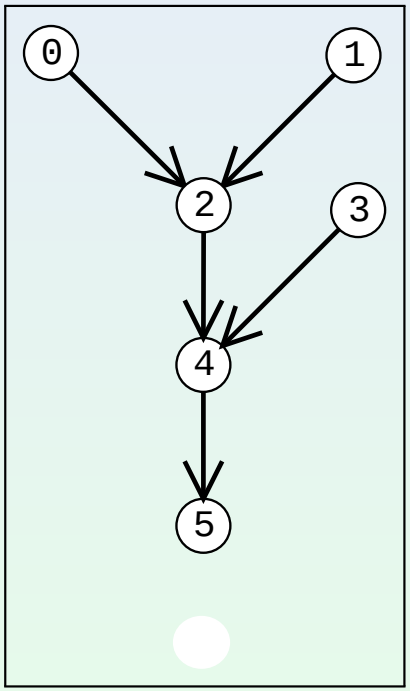
- The program can deal with gene perturbations in case of dynamic Bayesian Networks.


Gen2
Gen3
Gen1
Gen4
Gen5
class I
class I
class I
class I
class I
class I
class I
class I
class I
class I
class I
class I
class I
class I
class II
class III
class III
class III

Consensus network



Tests based on artificial networks



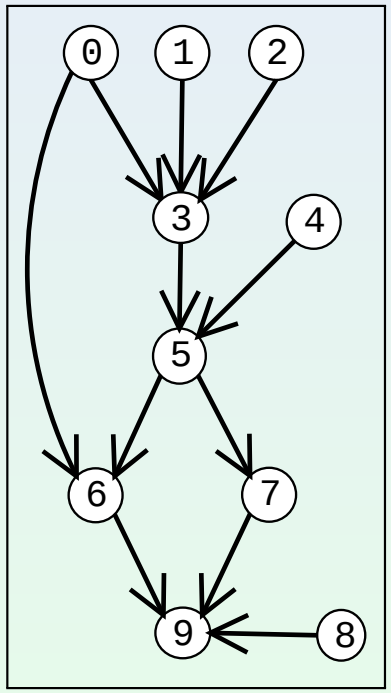
Network I

Description

- Each node (random variable) can take values 0 or 1.
- If a node has no parents it takes value 1 with the probability 0.5.
- Otherwise it takes value 1 with probability

$$\frac{\sum_{i=1}^{r} val(v_i)}{r},$$

where r is the number of parent nodes.



Network II

Number of networks in the optimal class

Results

Number of standard networks in the optimal class:

Network I	*6*	*12*	*36*	*60*	*120*		**Network II**	*10*	*20*	*100*	*200*
BDe	4	2	2	2	2		**BDe**	27	16	8	8
MDL	4	7	6	2	2		**MDL**	135	72	8	32

Number of dynamic networks in the optimal class:

Network I	*6*	*12*	*36*	*60*	*120*		**Network II**	*10*	*20*	*100*	*200*
BDe	1	1	1	1	1		**BDe**	4	1	1	1
MDL	1	1	1	1	1		**MDL**	4	1	1	1

Conclusions

- BDe score function can better distinguish between networks.
- The dynamic formalism significantly reduces the number of networks found in the optimal class.

Number of networks in the optimal class

Results

Number of standard networks in the optimal class:

Network I	*6*	*12*	*36*	*60*	*120*		**Network II**	*10*	*20*	*100*	*200*
BDe	4	2	2	2	2		**BDe**	27	16	8	8
MDL	4	7	6	2	2		**MDL**	135	72	8	32

Number of dynamic networks in the optimal class:

Network I	*6*	*12*	*36*	*60*	*120*		**Network II**	*10*	*20*	*100*	*200*
BDe	1	1	1	1	1		**BDe**	4	1	1	1
MDL	1	1	1	1	1		**MDL**	4	1	1	1

Conclusions

- BDe score function can better distinguish between networks.
- The dynamic formalism significantly reduces the number of networks found in the optimal class.

True-positive vs. false-positive edges

Results

Number of true-positive/number of false-positive edges found with standard Bayesian Networks

Network I	*6*	*12*	*36*	*60*	*120*	**Network II**	*10*	*20*	*100*	*200*
BDe	2 / 2	1 / 4	4 / 4	3 / 5	5 / 3	**BDe**	3 / 9	5 / 13	9 / 5	11 / 5
MDL	2 / 2	2 / 6	5 / 5	3 / 5	5 / 3	**MDL**	5 / 13	7 / 15	9 / 5	11 / 8

Number of true-positive/number of false-positive edges found with dynamic Bayesian Networks

Network I	*6*	*12*	*36*	*60*	*120*	**Network II**	*10*	*20*	*100*	*200*
BDe	1 / 0	2 / 0	4 / 1	5 / 0	5 / 0	**BDe**	3 / 2 *	3 / 1	9 / 0	11 / 0
MDL	1 / 0	2 / 0	4 / 1	5 / 0	5 / 0	**MDL**	3 / 2 *	3 / 1	9 / 0	11 / 0

Conclusions

If the set of expression data is large enough, the Dynamic Bayesian Networks formalism is capable of finding the source network.

True-positive vs. false-positive edges

Results

Number of true-positive/number of false-positive edges found with standard Bayesian Networks

Network I	*6*	*12*	*36*	*60*	*120*	**Network II**	*10*	*20*	*100*	*200*
BDe	2 / 2	1 / 4	4 / 4	3 / 5	5 / 3	**BDe**	3 / 9	5 / 13	9 / 5	11 / 5
MDL	2 / 2	2 / 6	5 / 5	3 / 5	5 / 3	**MDL**	5 / 13	7 / 15	9 / 5	11 / 8

Number of true-positive/number of false-positive edges found with dynamic Bayesian Networks

Network I	*6*	*12*	*36*	*60*	*120*	**Network II**	*10*	*20*	*100*	*200*
BDe	1 / 0	2 / 0	4 / 1	5 / 0	5 / 0	**BDe**	3 / 2 *	3 / 1	9 / 0	11 / 0
MDL	1 / 0	2 / 0	4 / 1	5 / 0	5 / 0	**MDL**	3 / 2 *	3 / 1	9 / 0	11 / 0

Conclusions

If the set of expression data is large enough, the Dynamic Bayesian Networks formalism is capable of finding the source network.

Bibliography

S. Ott, S. Imoto, and S. Miyano

Finding Optimal Models for Small Gene Networks.
Pacific Symposium on Biocomputing, 9:557-567, 2004.

Nir Friedman and Moises Goldszmidt

Learning Bayesian Networks with Local Structure.
Twelfth Conference on Uncertainty in Artificial Intelligence, 252-262, 1996.

Gregory F. Cooper and Edward Herskovits

A Bayesian Method for the Induction of Probabilistic Networks from Data.
Machine Learning, 9:309-347, 1992.

Hidde de Jong

Modeling and Simulation of Genetic Regulatory Systems: A Literature Review.
Journal of Computational Biology, 9:67-103, 2002.

S. Ott, A. Hansen, S.-Y. Kim, and S. Miyano

Superiority of network motifs over optimal networks and an application to the revelation of gene network evolution.
Bioinformatics, 21:227-238, 2005.

Norbert Dojer, Anna Gambin, Andrzej Mizera, Bartek Wilczynski, and Jerzy Tiuryn

Applying dynamic Bayesian networks to perturbed gene expression data.
BMC Bioinformatics, 7:249, 2006.

Bibliography

Nir Friedman, Michal Linial, Iftach Nachman, and Dana Pe'er

Using Bayesian networks to analyze expression data.
Journal of Computational Biology, 7:601-620, 2000.

Daniel E. Zak, Francis J. Doyle III, Gregory E. Gonye, and James S. Schwaber

Simulation studies for the identification of genetic networks from cDNA array and regulatory activity data.
Proceedings of the Second International Conference on Systems Biology, 231-238, 2001.

David Heckerman, Dan Geiger, and David M. Chickering

Learning Bayesian networks: The combination of knowledge and statistical data.
Machine Learning, 20:197-243, 1995.

Thomas Verma, and Judea Pearl

Equivalence and synthesis of causal models.
Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence, 220-227, 1990.

D. M. Chickering

Learning bayesian networks is NP-complete.
Learning from Data: Artificial Intelligence and Statistics V, Springer-Verlag, 1996.

J. Rissanen

Modelling by shortest data description.
Automatica, 14:465-471, 1978.

Thank you for your attention!