ulm university universität ulm

The *BoolNet* package

Hans A. Kestler

with Christoph Müssel

Medical Systems Biology
Ulm University, Germany

The *BoolNet* package

- R package for construction, analysis and simulation of Boolean networks
- Open-source and freely available at CRAN (cran.r-project.org)
- Main features:
  - supports synchronous, asynchronous and probabilistic Boolean networks
  - reconstruction of networks from time series
  - analysis of dynamic behaviour of networks (attractor search & Markov chain simulation)
  - visualization of static and dynamic network properties
  - generation of random networks & time series, hypothesis testing
  - import and export of several network formats (SBML qual, Biotapestry, Pajek)

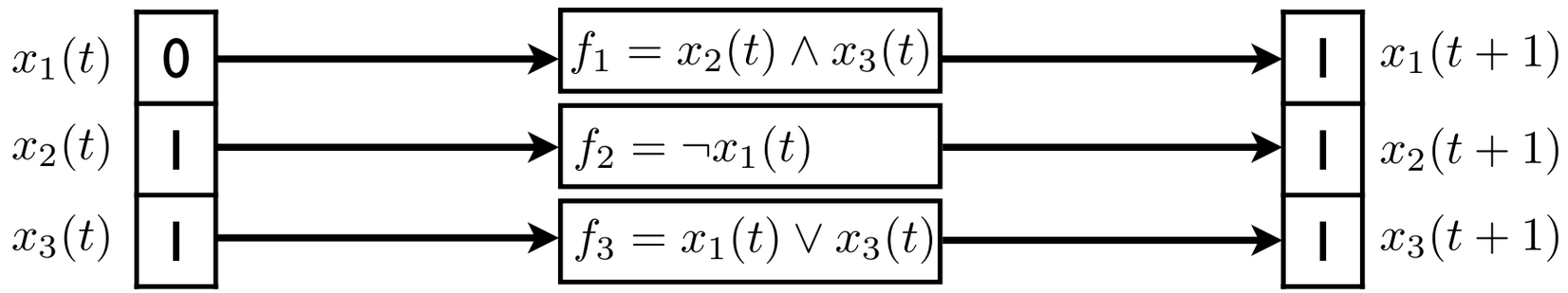
[1] Müssel, Hopfensitz, Kestler, Bioinformatics, 2010

Why R?

- Widely used in biostatistics & systems biology
- Interface to a high number of extension packages, particularly for these research fields
  (e.g., ~750 packages in the Bioconductor repository, ~5400 packages on CRAN)
- R and its packages are free and open-source
- Interactive prompt enables flexible combination of different commands
- Convenient functions to import and export various types of data
- Many preprocessing methods available (or being developed, see later)

Network types supported by *BoolNet*

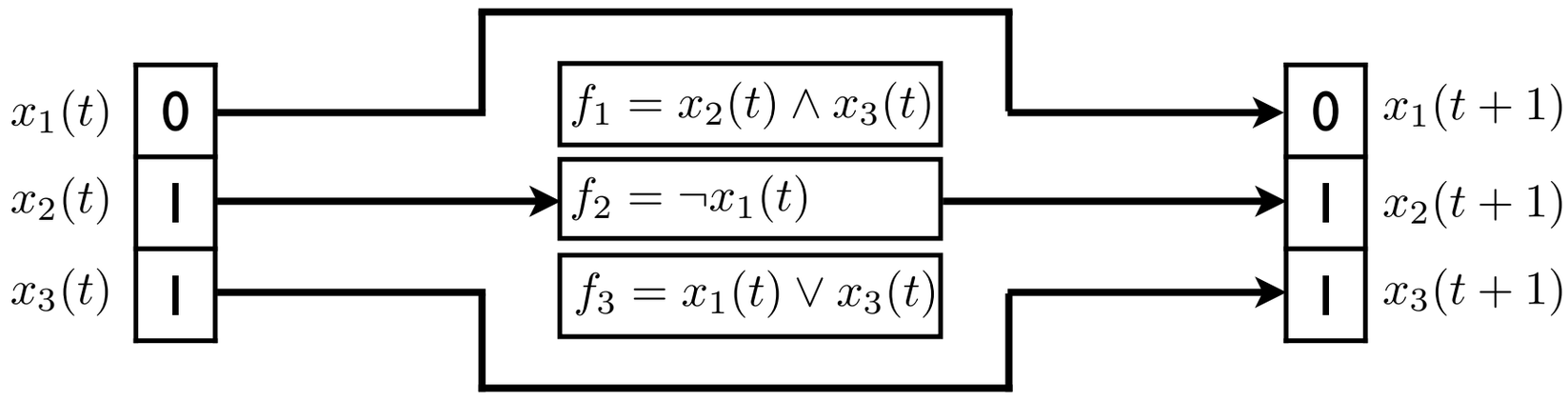
- **Synchronous Boolean networks [1]:**
  all genes updated synchronously



[1] Kauffman, J. Theor. Biol.,1969

Network types supported by *BoolNet*

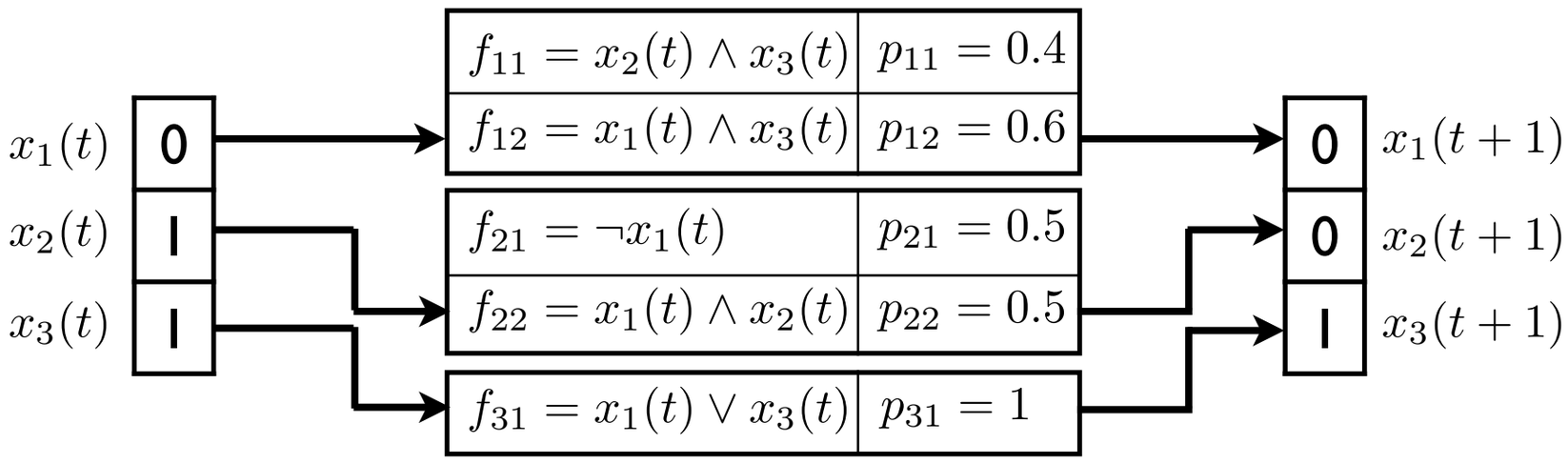
- **Asynchronous Boolean networks [2]:**
  one gene updated in each time step



[2] Thomas, J. Theor. Biol., 1991

Network types supported by *BoolNet*

- **Probabilistic Boolean networks [3]:**
  multiple transition functions per gene, random synchronous update

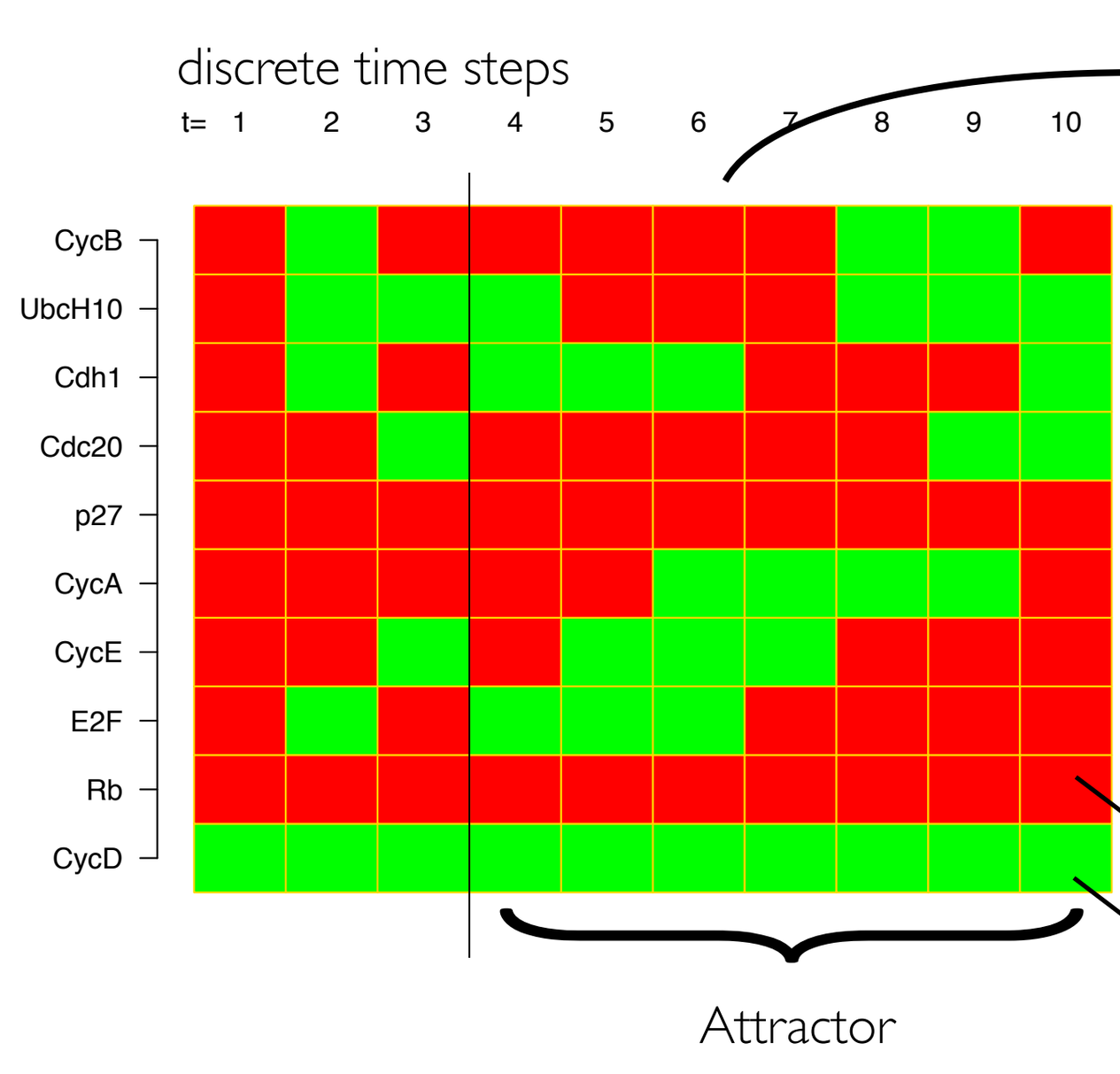


[3] Shmulevich et al., Bioinformatics, 2002

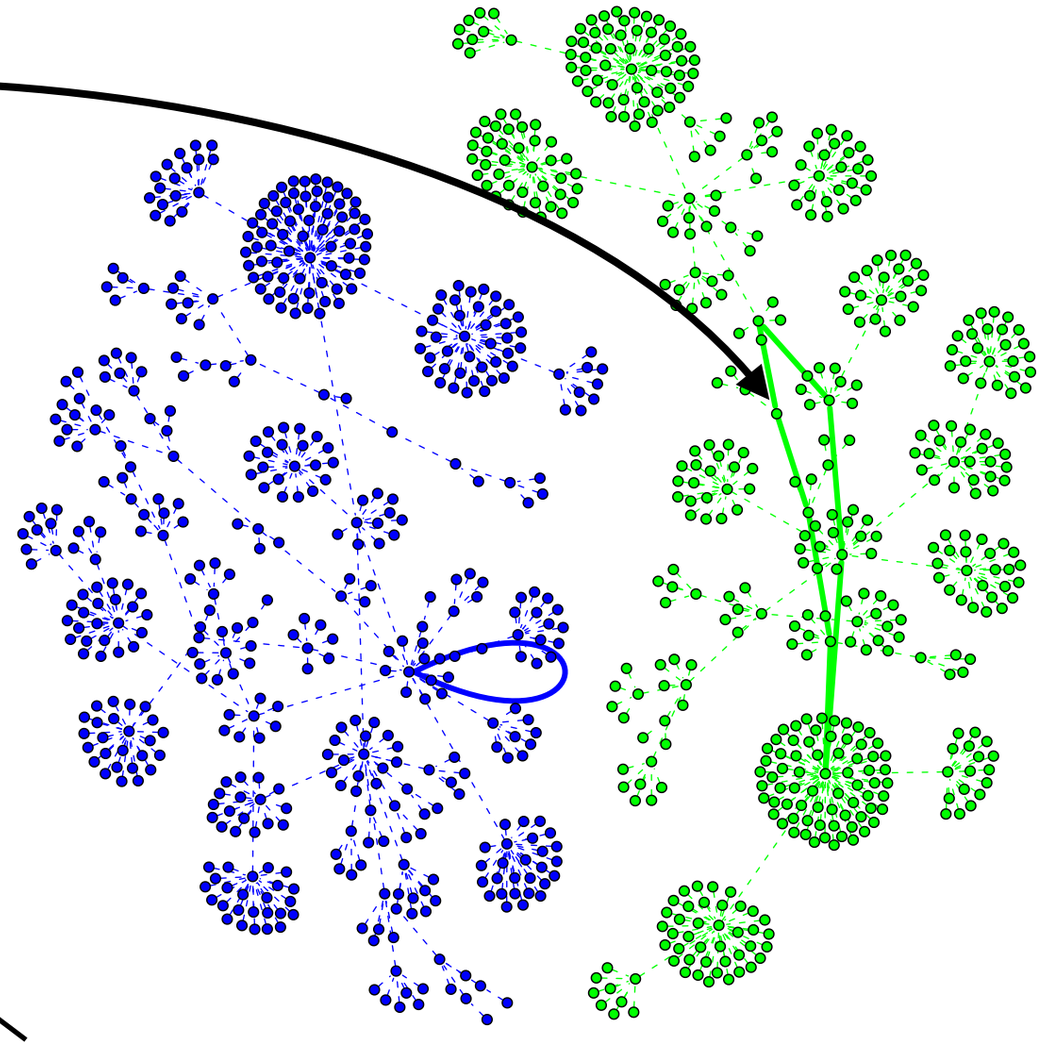
Dynamic analysis of Boolean networks

Simulation and attractor search for synchronous Boolean networks

Visualization of the 7 states of a cycle attractor:



State transitions and basins of attraction:

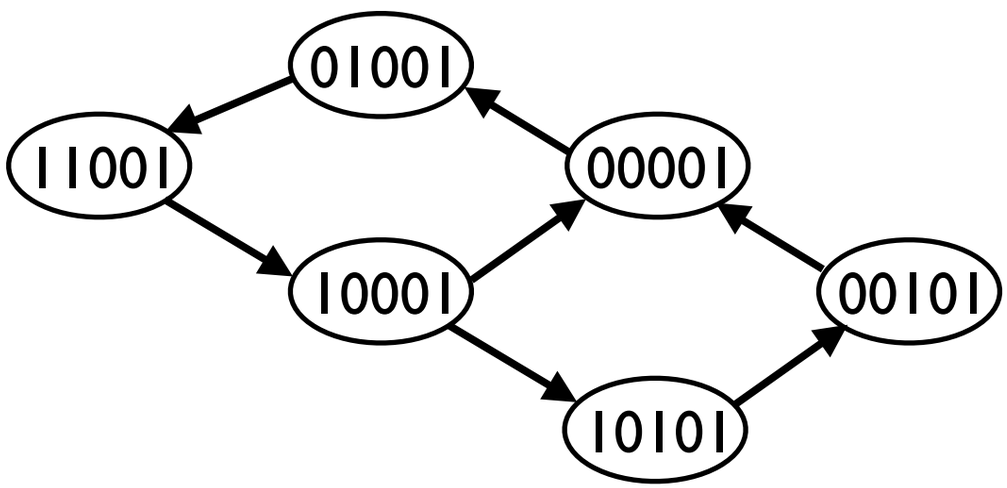


Mammalian cell cycle network, Fauré, Naldi, Chaouiya, Thieffry, Bioinformatics, 2006

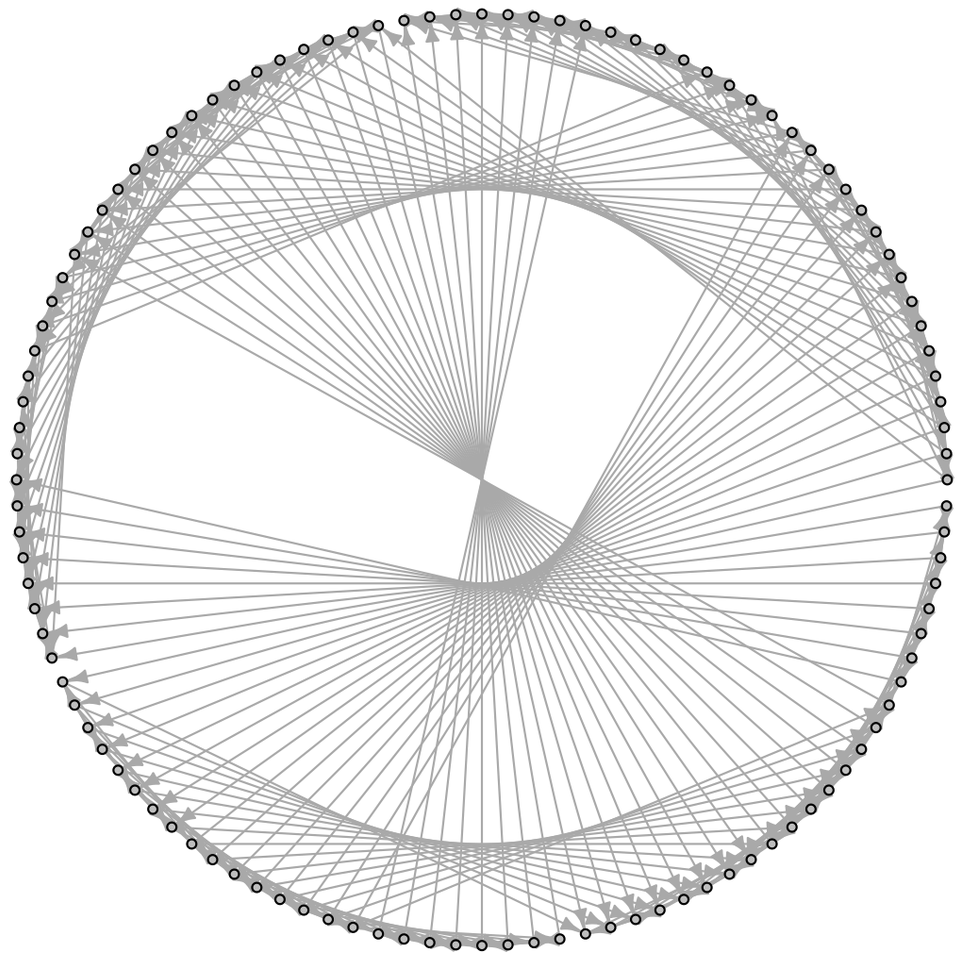
Dynamic analysis of Boolean networks

Identification of complex attractors in asynchronous Boolean networks

Complex attractors: all attractor states reachable from all other attractor states



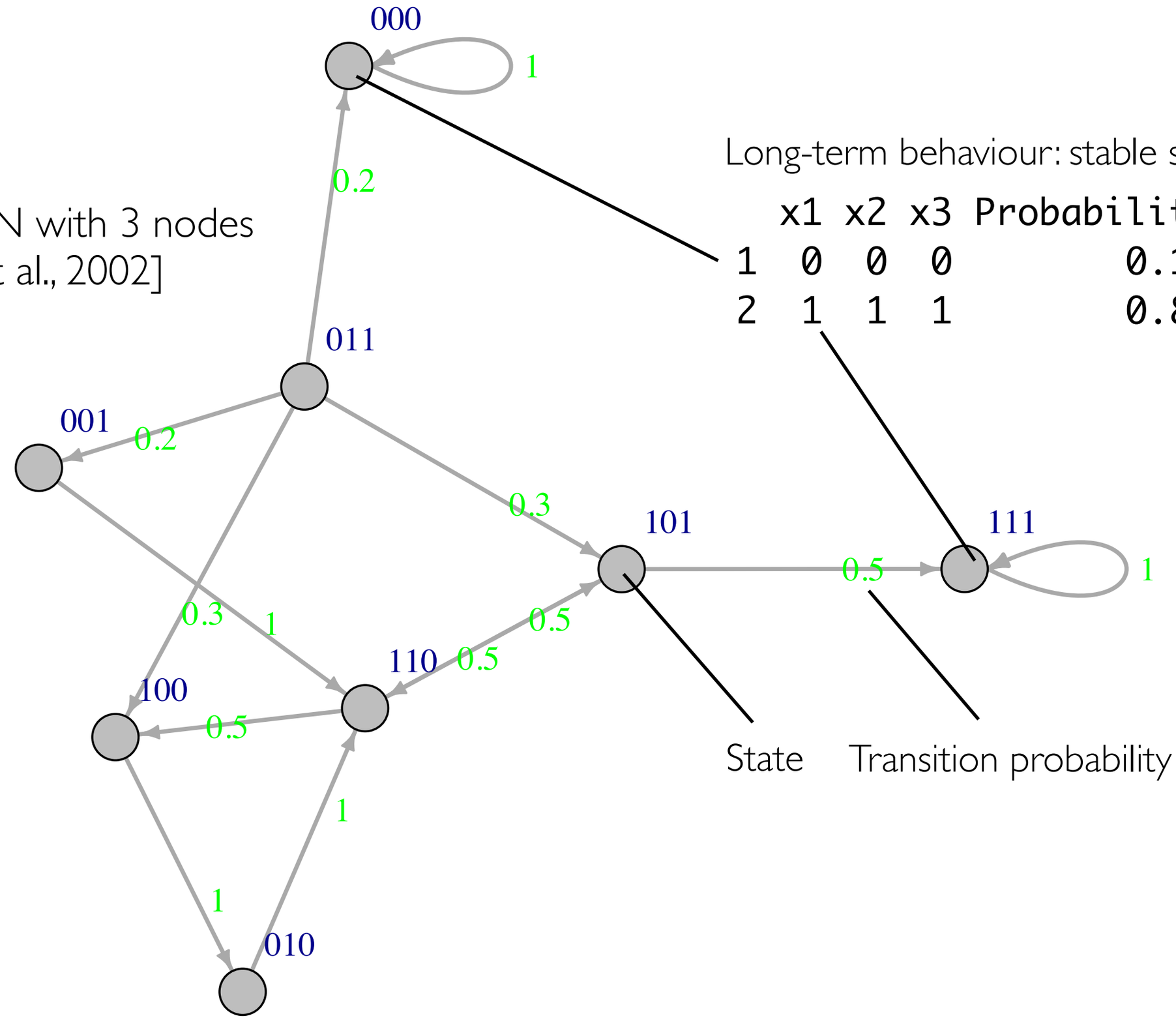
Visualization of an asynchronous attractor in a cell cycle network model:



Dynamic analysis of Boolean networks

Markov chain simulations for probabilistic Boolean networks

Simulation of
exemplary PBN with 3 nodes
[Shmulevich et al., 2002]



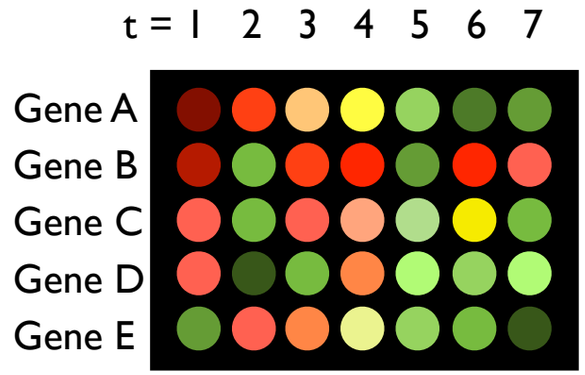
Long-term behaviour: stable states

	x1	x2	x3	Probability
1	0	0	0	0.15
2	1	1	1	0.85

State Transition probability

Reverse-engineering of Boolean networks

- *BoolNet* includes two well-known algorithms for the inference of Boolean networks from time series:
  - REVEAL [1]
  - Best-fit extension [2]
- To convert real-valued measurements to binary vectors, simple binarization techniques are included.



$\in \mathbb{R}^{n \times T}$

Binarization

$$\begin{pmatrix} 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} \in \mathbb{B}^{n \times T}$$

Reverse-engineering

$$\begin{aligned} Gene_1(t+1) &= Gene_1(t) \vee Gene_3(t) \\ Gene_2(t+1) &= \neg Gene_3 \wedge Gene_5(t) \\ Gene_3(t+1) &= Gene_5(t) \\ Gene_4(t+1) &= (Gene_3(t) \wedge Gene_2(t)) \vee Gene_5(t) \\ Gene_5(t+1) &= Gene_1(t) \end{aligned}$$

[1] Liang et al, Pacific Symposium on Biocomputing, 1998
[2] Lähdesmäki et al, Machine Learning, 2003

Random Boolean networks

- Random Boolean networks have been proposed to study the overall behaviour of certain network types and to identify specific properties of biological systems [1,2]
- *BoolNet* includes a generic facility for the generation of random Boolean networks with different topological and dynamic properties
- Network models can be tested against populations of random networks to evaluate their biological plausibility

[1] Kauffman, J. Theor. Biol.,1969
[2] Kauffman, Oxford University Press, 1993

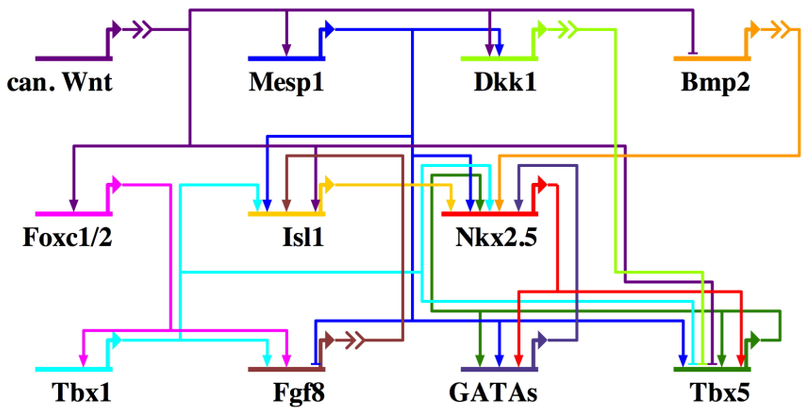
Import file formats

BoolNet can import from various formats:

- *BoolNet* network rule format (plain-text CSV, all types of networks):

```
targets, factors
p53, ATM & !Mdm2
Mdm2, !p53
```

- BioTapestry (synchronous and asynchronous Boolean networks)

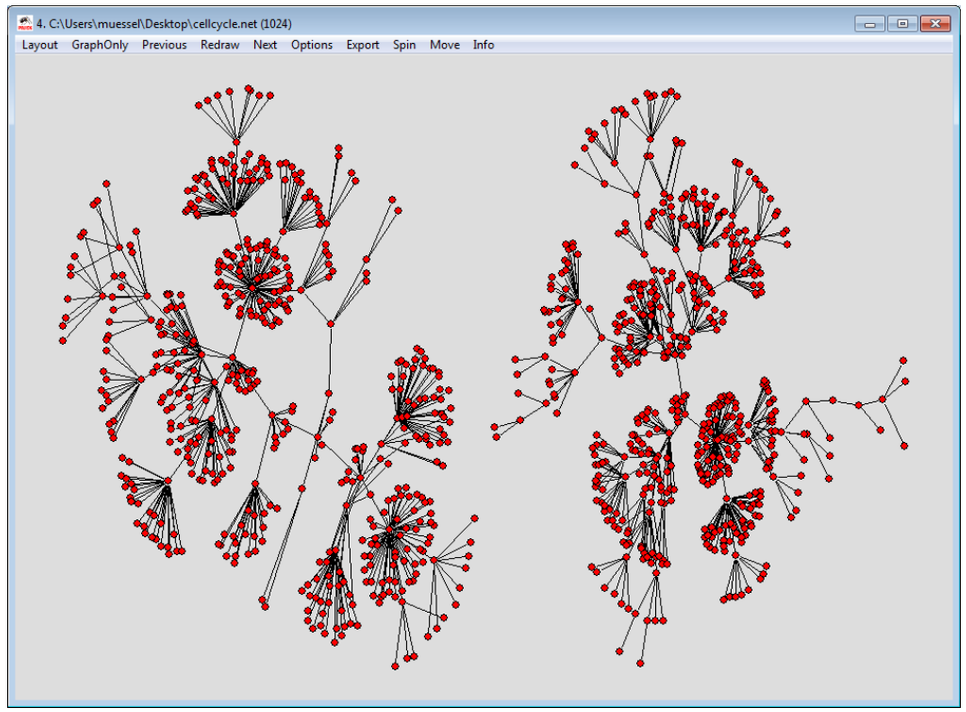


- SBML qual (synchronous and asynchronous Boolean networks)
  - General XML-based parser understanding the subset of SBML qual that describes logical models with binary genes only

Export file formats

Furthermore, different export formats are available:

- *BoolNet* network rule format
- Pajek (state transition graphs)

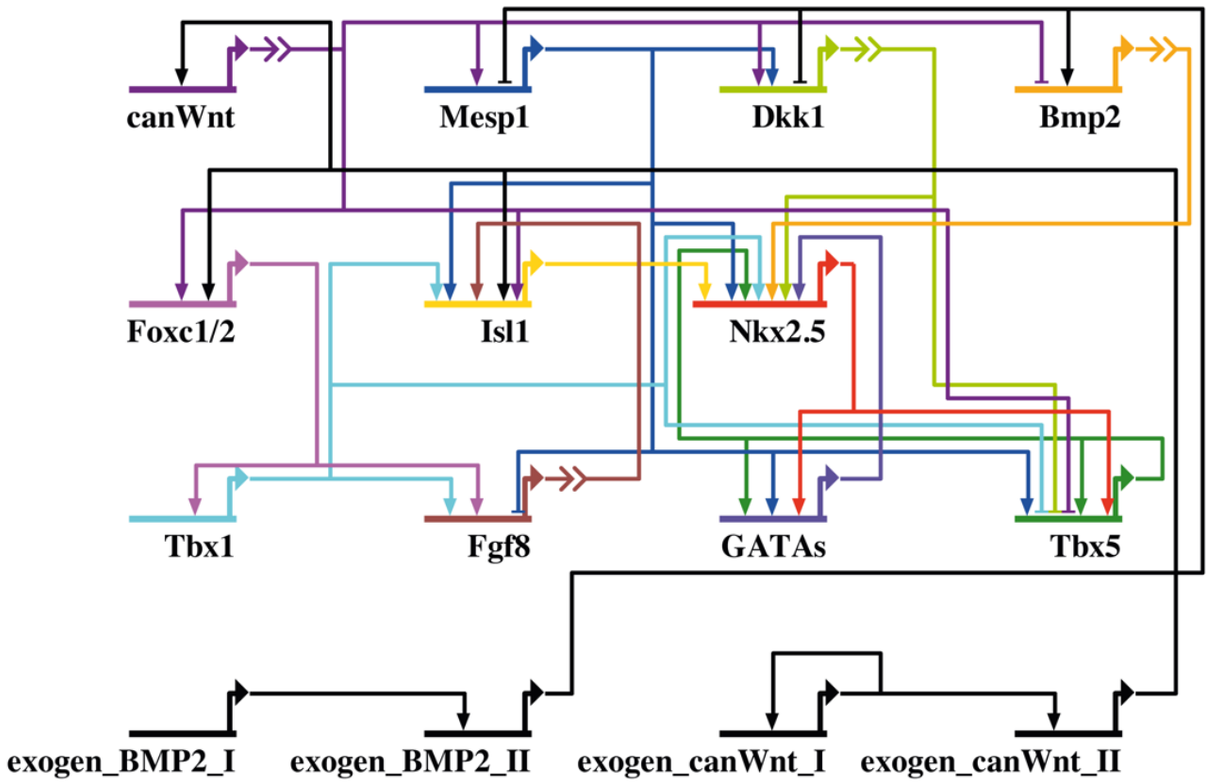


- SBML qual
  - Fully compliant with SBML validator and RelaxNG schema
  - Export and re-import of SBML in *BoolNet* is lossless

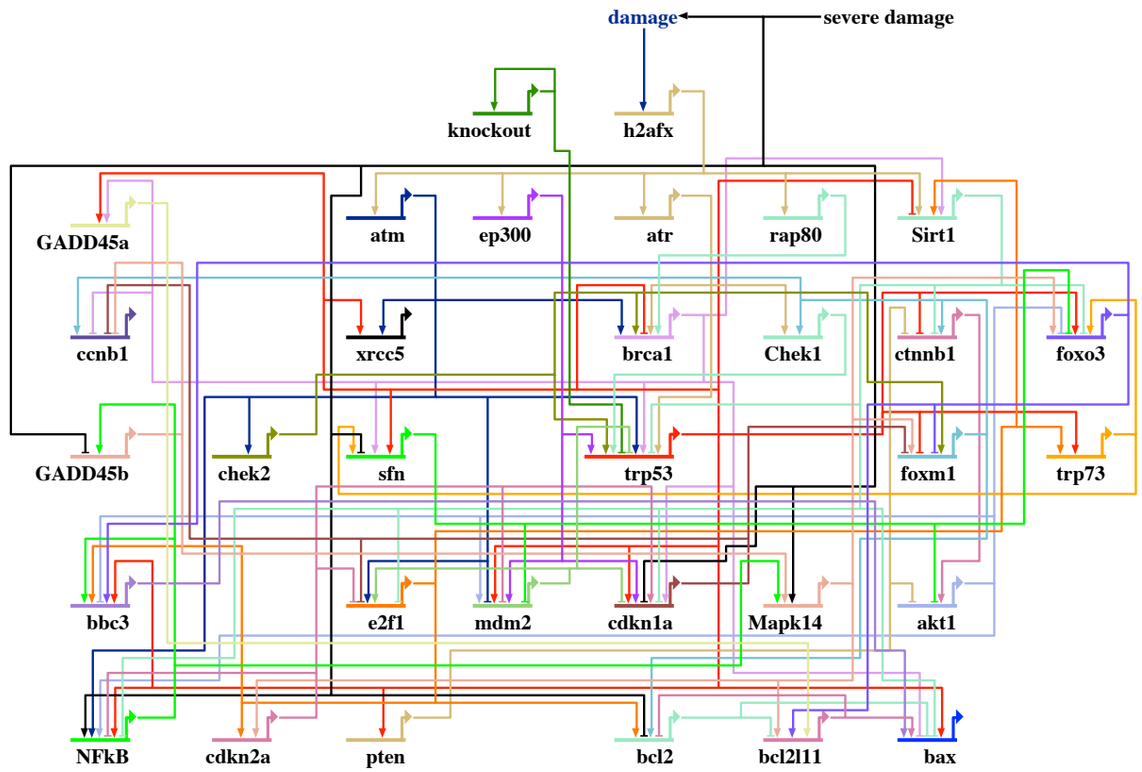
Applications

Several Boolean models have been/are currently being developed in our group:

Murine cardiac development [1]:
Models early heart development (first/second heart field) using 11 genes and 4 inputs



DNA damage response:
Models the p53 signaling pathway using 31 genes and 3 inputs

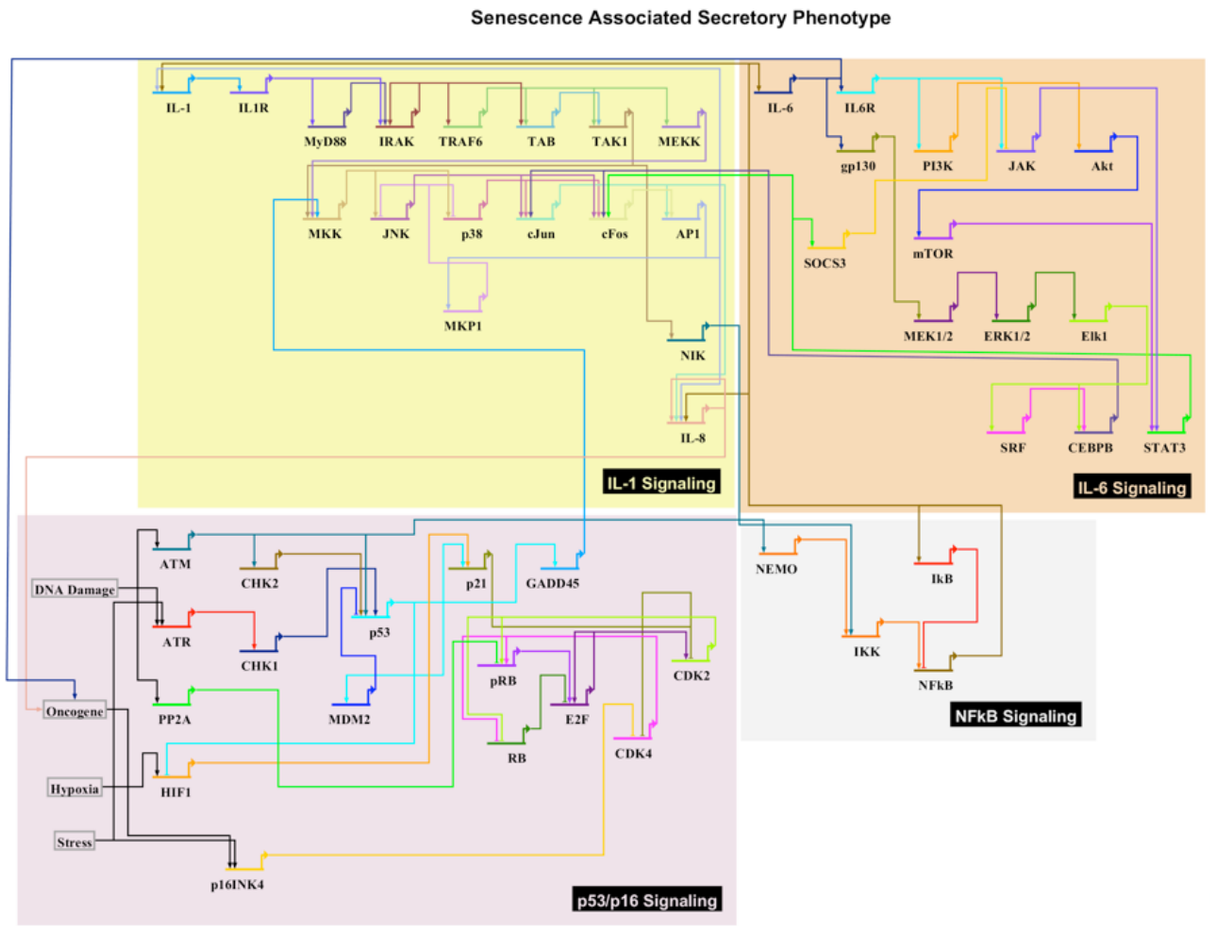


[1] Herrmann et al., PLOS ONE, 2012

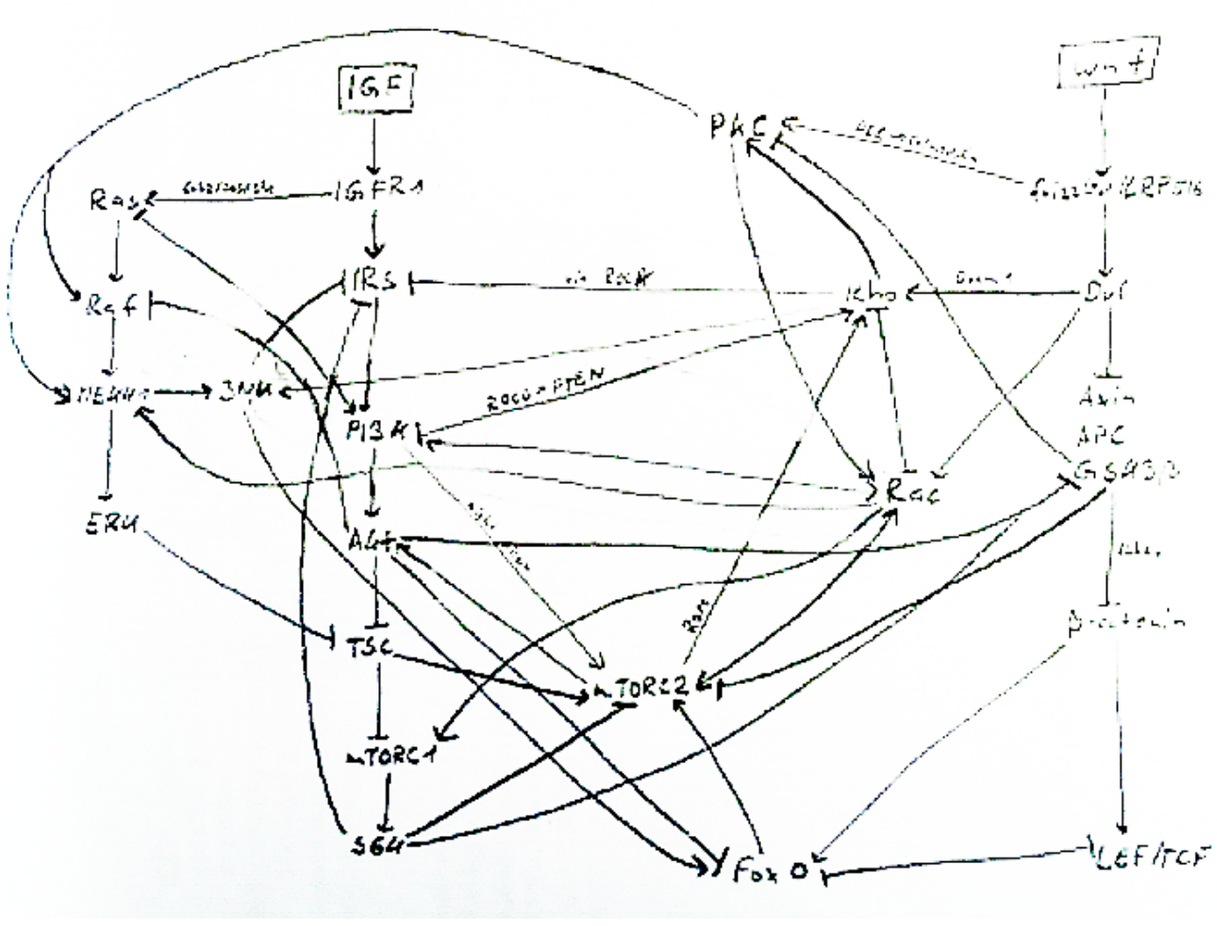
Applications

Several Boolean models have been/are currently being developed in our group:

Senescence-associated secretory phenotype: Models NFκB/IL-1/IL-6 signaling using 52 genes and 3 inputs



Interactions between Wnt signaling and IGF signaling (currently in an early stage):



Future developments I

New features are currently being integrated:

Extension of the supported model classes:

- support for temporal statements and delays
- additional operators (e.g. majority vote)
- implementation in a new simulator that is based on symbolic expression trees instead of truth tables

Examples:

- Gene A is activated two time steps af
  `A, C[-2]`
- Gene A is active if Gene B or Gene
  last three time steps:
  `A, any[t=-3..-1](B[t] | C[t]`
- Gene A is active if the majority of its
  present in the last two time steps
  `A, all[t=-2..-1](maj(B[t],C[`
- Gene A is inactivated after the first five time steps:
  `A, timelt(6)`

SBML qual specification:
Since the concept of time is beyond the scope of this specification it is recommended that the csymbols "time" and "delay" that explicitly involve time are not used.

➡ Support of time delays in SBML qual would be useful to be able to exchange such models

Future developments II

Extension of the integrated reconstruction methods:

- inclusion of prior knowledge, e.g.
  - known regulatory dependencies
  - dependencies identified by static methods
    e.g. identification of regulators by higher-order correlations [1,2]
- direct support for perturbation data
  (systematic knock-out/overexpression of upstream regulators)

Extension of the random network generator:

- integration of custom generation functions for specific function classes, e.g. canalyzing / nested canalyzing functions

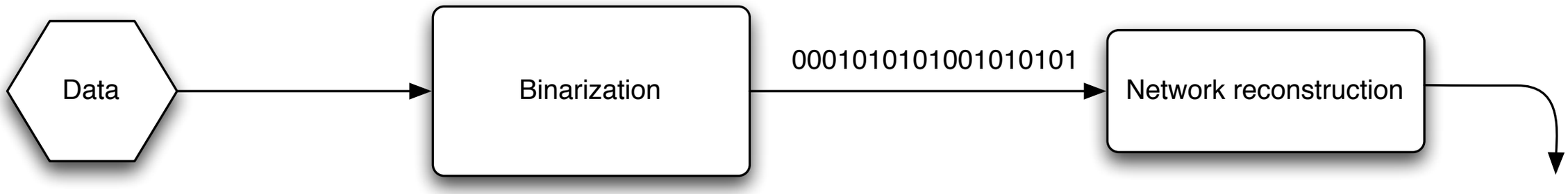
➡ **If you are interested in testing the new package (BoolNet 2.0), feel free to ask!**

[1] Maucher et al., Computational Statistics, 2014
[2] Maucher et al., Bioinformatics, 2011

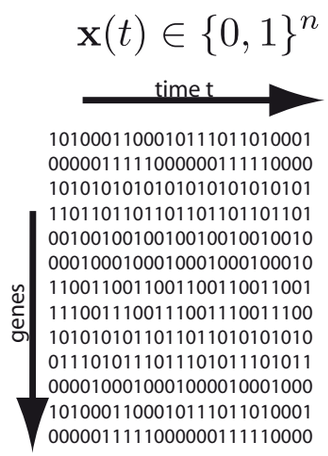
Future developments III - R package binarize

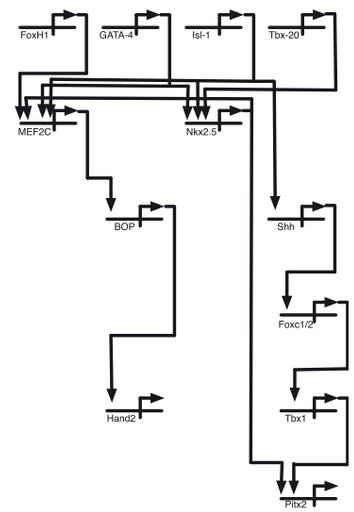
Reducing search complexity via robust binarization

Conventional approach



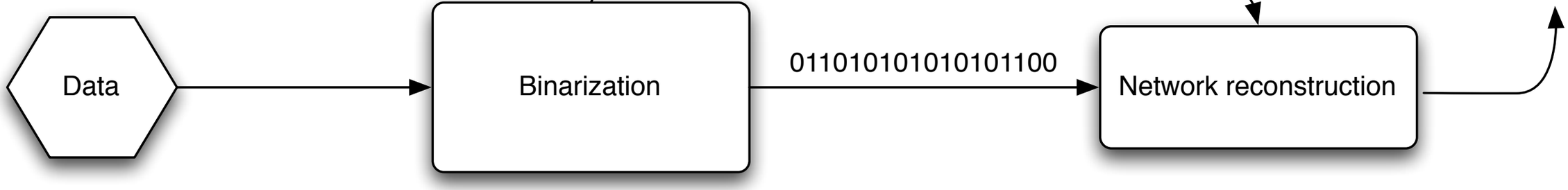
Expression profiles





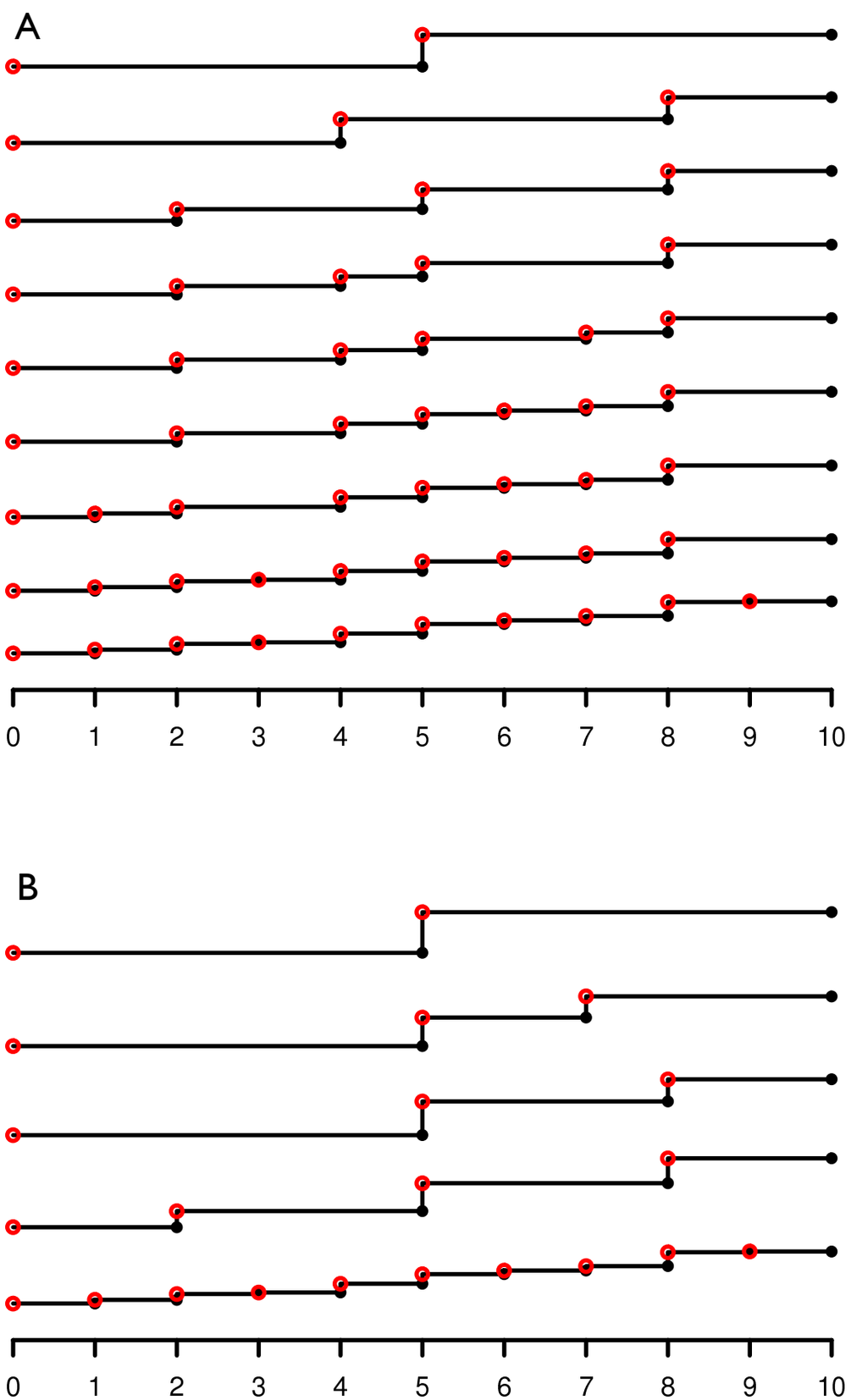
BASC

Feature reduction via validity of binarization



[Hopfensitz et al., IEEE/ACM T Comp Biol, & Bioinf, 2012]

Future developments III - R package binarize



Binarization Across multiple SCales (BASC)

Procedure

sorted real values

compute a sequence of step functions

optimal step functions

find the strongest single discontinuity in each step function

positions of the best steps

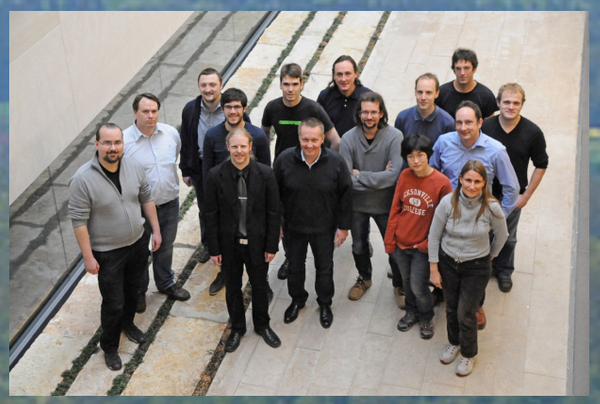
estimate location and variation of the strongest discontinuities

threshold, p-value, binarized data

family of 1-D time series
A: minimizing Euclidean distance to original
B: scale space approximation

[Hopfensitz et al., IEEE/ACM T Comp Biol, & Bioinf, 2012]

Bright views for logical modeling



Medical Systems Biology, Ulm University

http://sysbio.uni-ulm.de

DFG Deutsche Forschungsgemeinschaft

SYSTAR

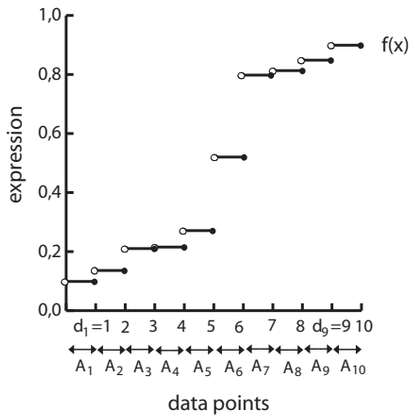
Federal Ministry of Education and Research

Binarization scenarios

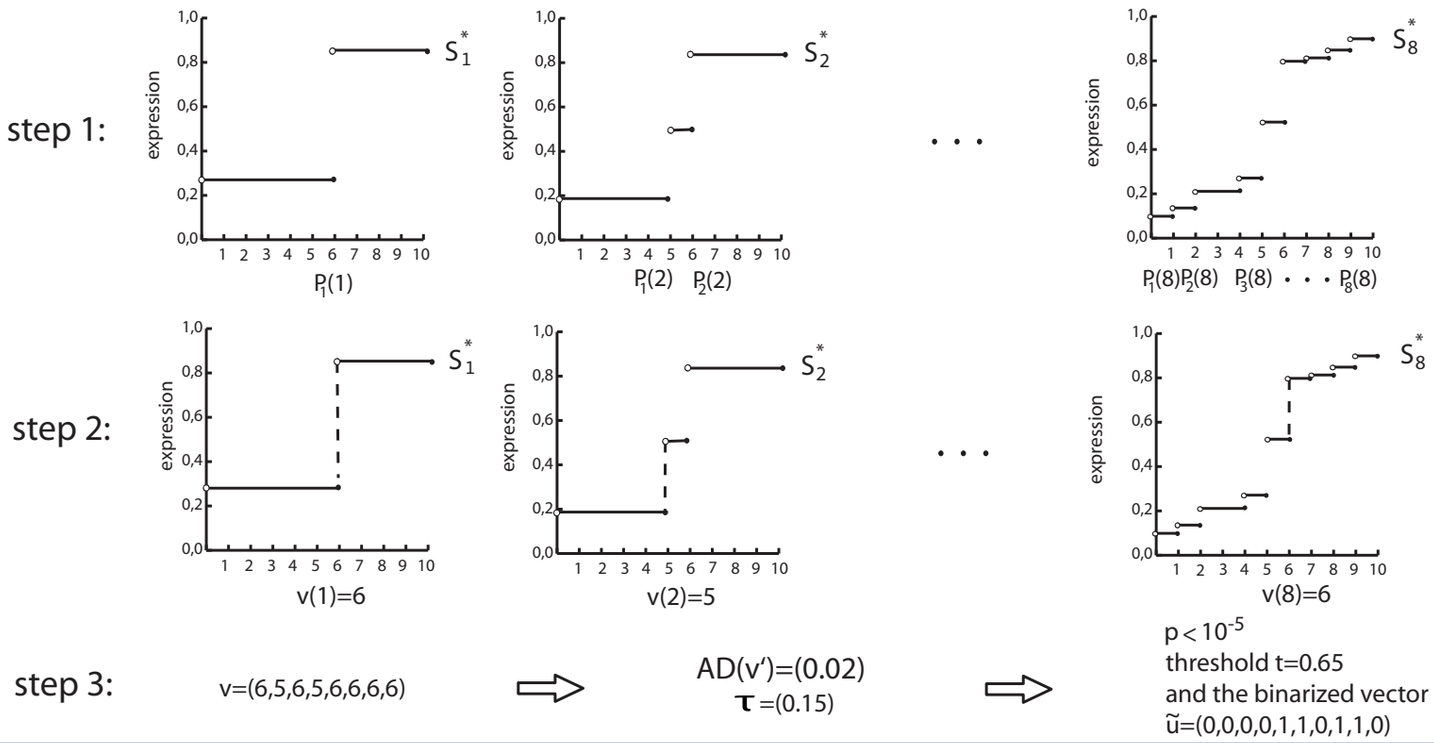
Scenario A

input vector:

u=(0.22, 0.29, 0.1, 0.13, 0.8, 0.9, 0.22, 0.85, 0.81, 0.5)



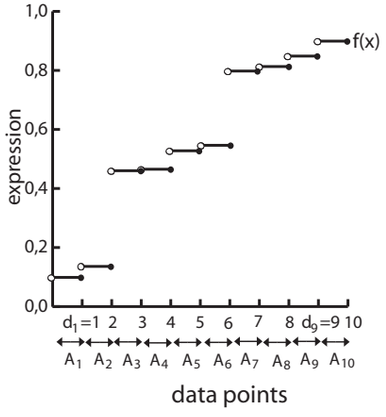
A: binarization



Scenario B

input vector:

u=(0.5, 0.57, 0.1, 0.13, 0.8, 0.9, 0.5, 0.85, 0.81, 0.58)



B: binarization