Bayesian Networks in Computational Life Science

Why use Bayesian Networks in Computational Life Science ?

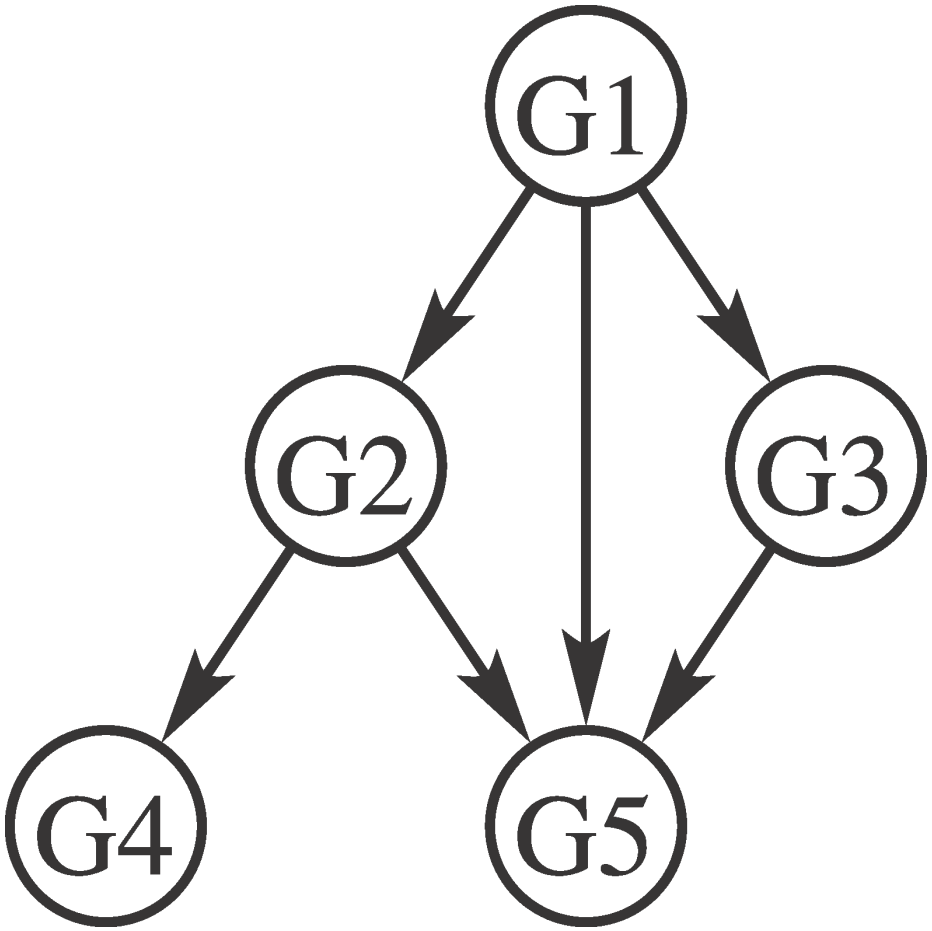
Low sample size: Marginalization

Incomplete data set: EM algorithm

Inherent noise: probabilistic estimates

Biological Knowledge: Prior Information

Picture adapted from “Needham, Chris J., et al. (2007)”

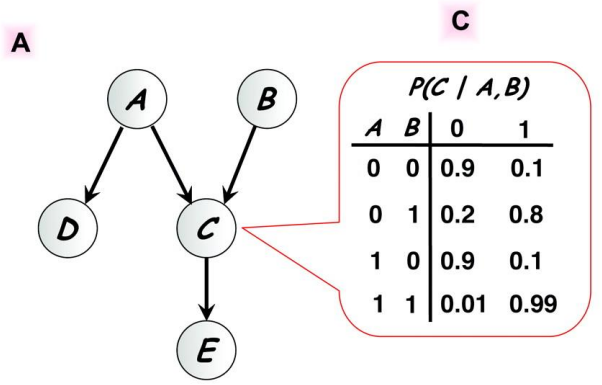


What are Bayesian Networks used for?

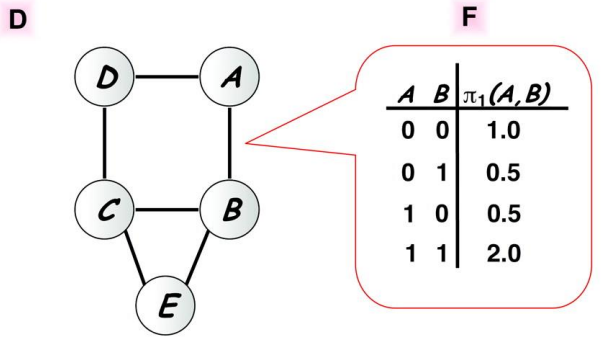
Probability distributions (parameter learning) : given some nodes (evidence) learn about the values for other nodes

Structure learning : given evidence on certain nodes learn structure of DAG

Picture adapted from “Friedman, Nir. "Inferring cellular networks using probabilistic graphical models." Science 303.5659 (2004): 799-805.”



B $P(A,B,C,D,E) = P(A)P(B)P(C \mid A,B)P(D \mid A)P(E \mid C)$



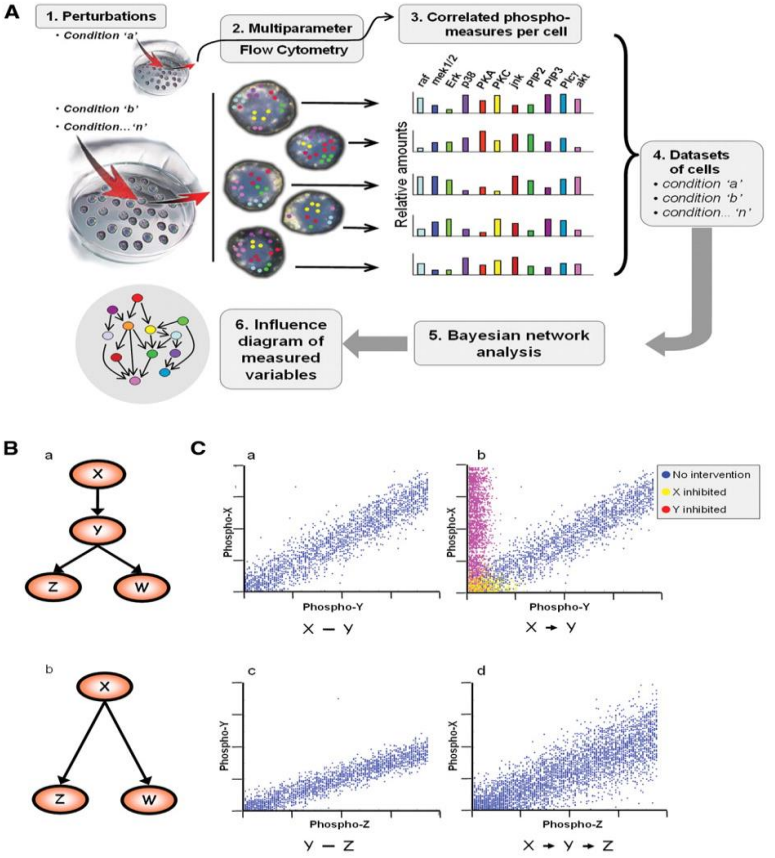
E $P(A,B,C,D,E) = \frac{1}{Z}\pi_1(A,B)\pi_2(B,C,E)\pi_3(C,D)\pi_4(A,D)$

Bayesian Networks in Computational Life sciences

Inferring Cellular Networks (Reverse Engineering)

Biological Data Integration

Learning causality (activation, inhibition) in protein signaling, gene regulatory networks -(refer fig.)



Picture adapted from “Sachs, Karen, et al. "Causal protein-signaling networks derived from multiparameter single-cell data." Science 308.5721 (2005): 523-529.”

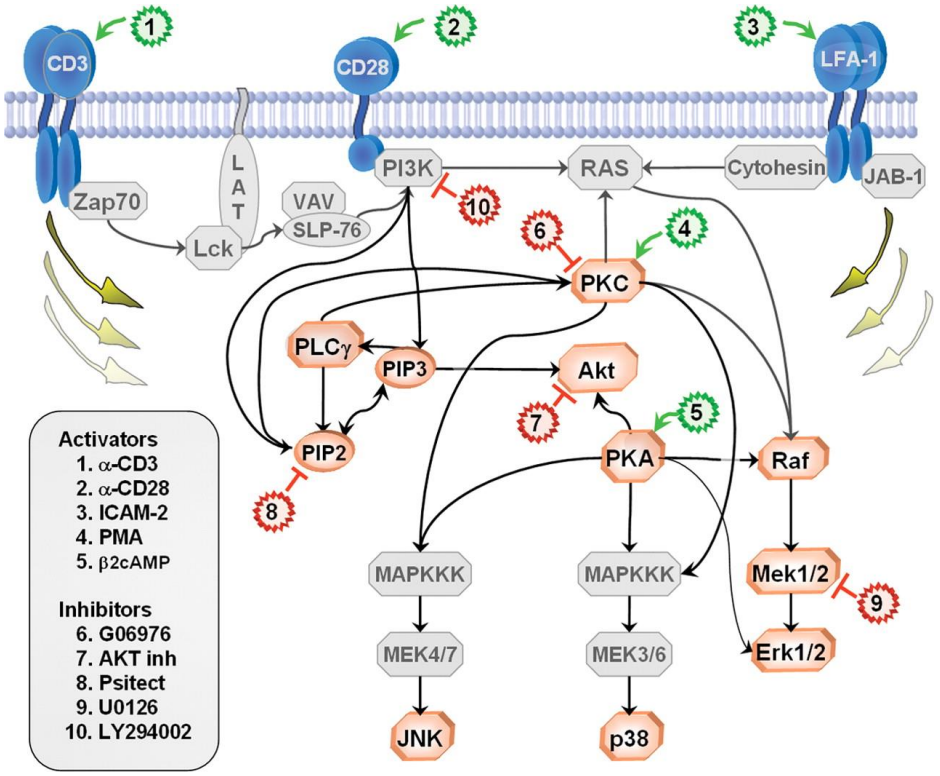
Motivation behind using Bayesian networks

Clustering of Gene Expression traditionally used to discover modules

Co-expression (or correlation) does not imply causality

Bayesian networks alternative to clustering in gene expression.

Causality requires knock-out experiments.



Picture adapted from “Sachs, Karen, et al. "Causal protein-signaling networks derived from multiparameter single-cell data." Science 308.5721 (2005): 523-529.”

Relationship between ODE based modelling

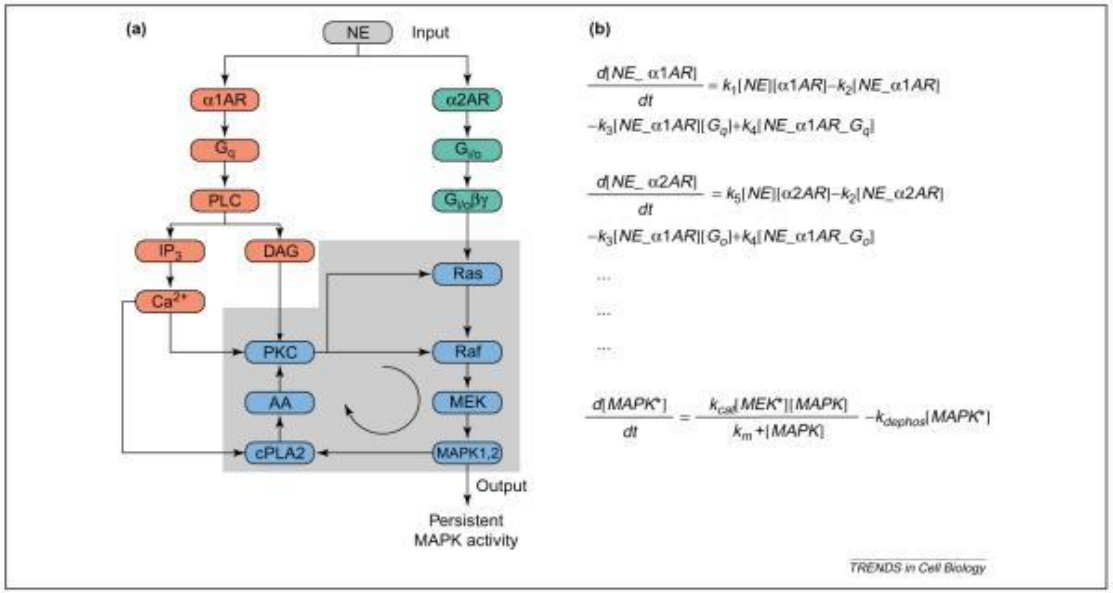
ODE models causality by directly assuming the form of the regulation:

$$\dot{x}_i(t) = f_i(x_1, \ldots, x_N, u, \theta_i)$$

Reverse Engineering means to infer parameters and function.

A lot of mechanistic knowledge needed as structure of the network is assumed to be known.

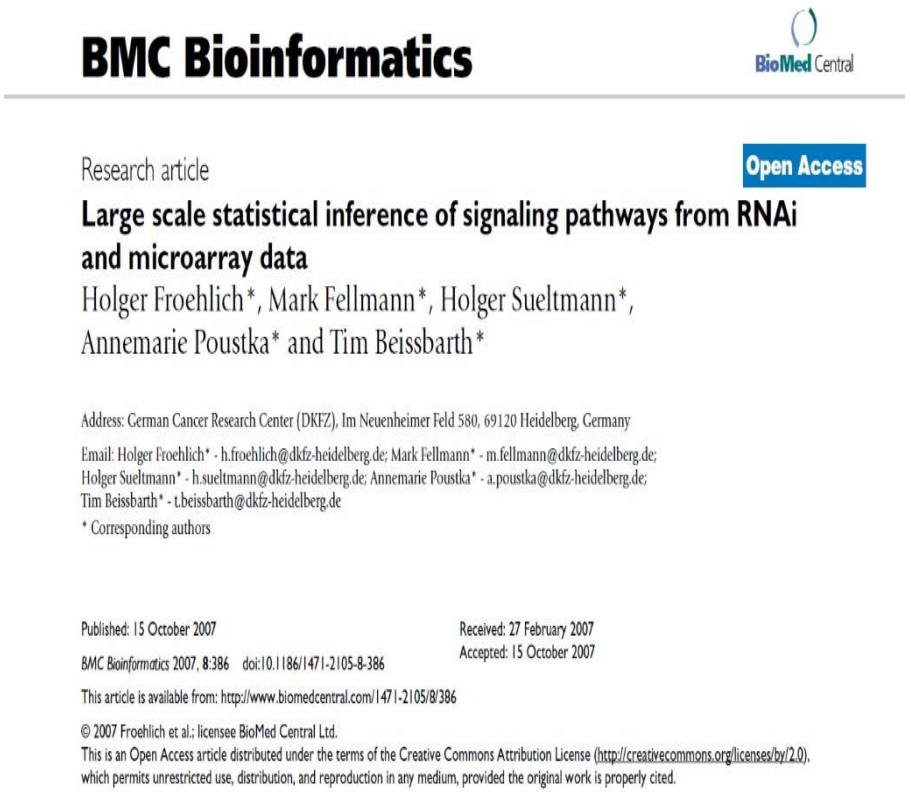
Ground truth for testing Bayesian Networks can be generated using known ODE systems



Data sets (from Bansal et al. 2007)

ID	Cell/organism	Type	Samples	Genes	Reference	True network
A	HumanBcells	S	254	7907	(Basso *et al*, 2005)	Twenty-six Myc targets (Basso *et al*, 2005)
B	*S. cerevisiae*	S	300	6312	(Hughes *et al*, 2000)	Eight hundred and forty-four TF–gene interactions (Lee *et al*, 2002)
C	HumanBcells	S	254	23	(Basso *et al*, 2005)	11 Myc targets + 11 non-targets (Basso *et al*, 2005)
D	*S. cerevisiae*	S	300	90	(Hughes *et al*, 2000)	Subset of TF–gene interactions (Lee *et al*, 2002)
E	*E. coli*	S	9	9	(Gardner *et al*, 2003)	Nine-gene network (Gardner *et al*, 2003)
F	*E. coli*	T	6	9	gardnerlab.bu.edu	Nine-gene network (Gardner *et al*, 2003)

Selected works on Bayesian Networks at ABI



BMC Bioinformatics

BioMed Central

Research article

Open Access

Large scale statistical inference of signaling pathways from RNAi and microarray data

Holger Froehlich*, Mark Fellmann*, Holger Sueltmann*, Annemarie Poustka* and Tim Beissbarth*

Address: German Cancer Research Center (DKFZ), Im Neuenheimer Feld 580, 69120 Heidelberg, Germany

Email: Holger Froehlich* - h.froehlich@dkfz-heidelberg.de; Mark Fellmann* - m.fellmann@dkfz-heidelberg.de; Holger Sueltmann* - h.sueltmann@dkfz-heidelberg.de; Annemarie Poustka* - a.poustka@dkfz-heidelberg.de; Tim Beissbarth* - t.beissbarth@dkfz-heidelberg.de

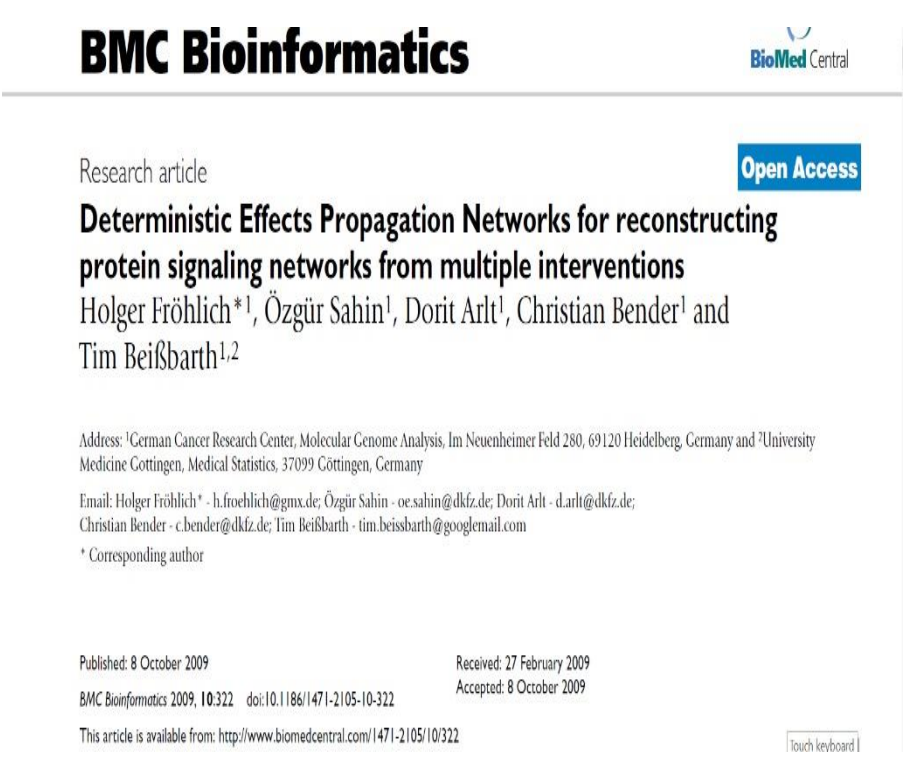
* Corresponding authors

Published: 15 October 2007

BMC Bioinformatics 2007, **8**:386 doi:10.1186/1471-2105-8-386

Received: 27 February 2007
Accepted: 15 October 2007

This article is available from: http://www.biomedcentral.com/1471-2105/8/386



BMC Bioinformatics

BioMed Central

Research article

Open Access

Deterministic Effects Propagation Networks for reconstructing protein signaling networks from multiple interventions

Holger Fröhlich*[1], Özgür Sahin[1], Dorit Arlt[1], Christian Bender[1] and Tim Beißbarth[1,2]

Address: [1]German Cancer Research Center, Molecular Genome Analysis, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany and [2]University Medicine Göttingen, Medical Statistics, 37099 Göttingen, Germany

Email: Holger Fröhlich* - h.froehlich@gmx.de; Özgür Sahin - oe.sahin@dkfz.de; Dorit Arlt - d.arlt@dkfz.de; Christian Bender - c.bender@dkfz.de; Tim Beißbarth - tim.beissbarth@googlemail.com

* Corresponding author

Published: 8 October 2009

BMC Bioinformatics 2009, **10**:322 doi:10.1186/1471-2105-10-322

Received: 27 February 2009
Accepted: 8 October 2009

This article is available from: http://www.biomedcentral.com/1471-2105/10/322

Selected works on Bayesian networks at ABI

BIOINFORMATICS

Vol. 26 ECCB 2010, pages i596–i602
doi:10.1093/bioinformatics/btq385

Dynamic deterministic effects propagation networks: learning signalling pathways from longitudinal protein array data

Christian Bender[1,*], Frauke Henjes[1], Holger Fröhlich[2], Stefan Wiemann[1], Ulrike Korf[1] and Tim Beißbarth[3]

[1]Department of Molecular Genome Analysis, German Cancer Research Center, 69120 Heidelberg, [2]Department of Algorithmic Bioinformatics, Bonn-Aachen International Center for IT, 53113 Bonn and [3]Department of Medical Statistics, University of Göttingen, 37099 Göttingen, Germany

ABSTRACT

Motivation: Network modelling in systems biology has become an important tool to study molecular interactions in cancer research, because understanding the interplay of proteins is necessary for developing novel drugs and therapies. *De novo* reconstruction of signalling pathways from data allows to unravel interactions between proteins and make qualitative statements on possible aberrations

Bayesian Networks (BN; Heckerman, 1996) have been frequently used to reconstruct gene regulatory networks from RNA expression experiments (Friedman *et al.*, 2000; Segal *et al.*, 2005) as well as causal protein–protein relationships for intensity data from protein quantification (Sachs *et al.*, 2005). The latter is an example where directed perturbations of several measured proteins were performed in order to resolve the structure of the underlying interactions.

BIOINFORMATICS **ORIGINAL PAPER**

Vol. 27 no. 2 2011, pages 238–244
doi:10.1093/bioinformatics/btq631

Systems biology

Advance Access publication November 10, 2010

Fast and efficient dynamic nested effects models

Holger Fröhlich[1,*], Paurush Praveen[1] and Achim Tresch[2]

[1]Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn-Aachen International Center for IT, Dahlmannstrasse 2, 53113 Bonn and [2]Department of Chemistry and Biochemistry, Ludwig-Maximilians-Universität München, Gene Center Munich and Center for integrated Protein Science CiPSM, Feodor-Lynen-Strasse 25, 81377 Munich, Germany

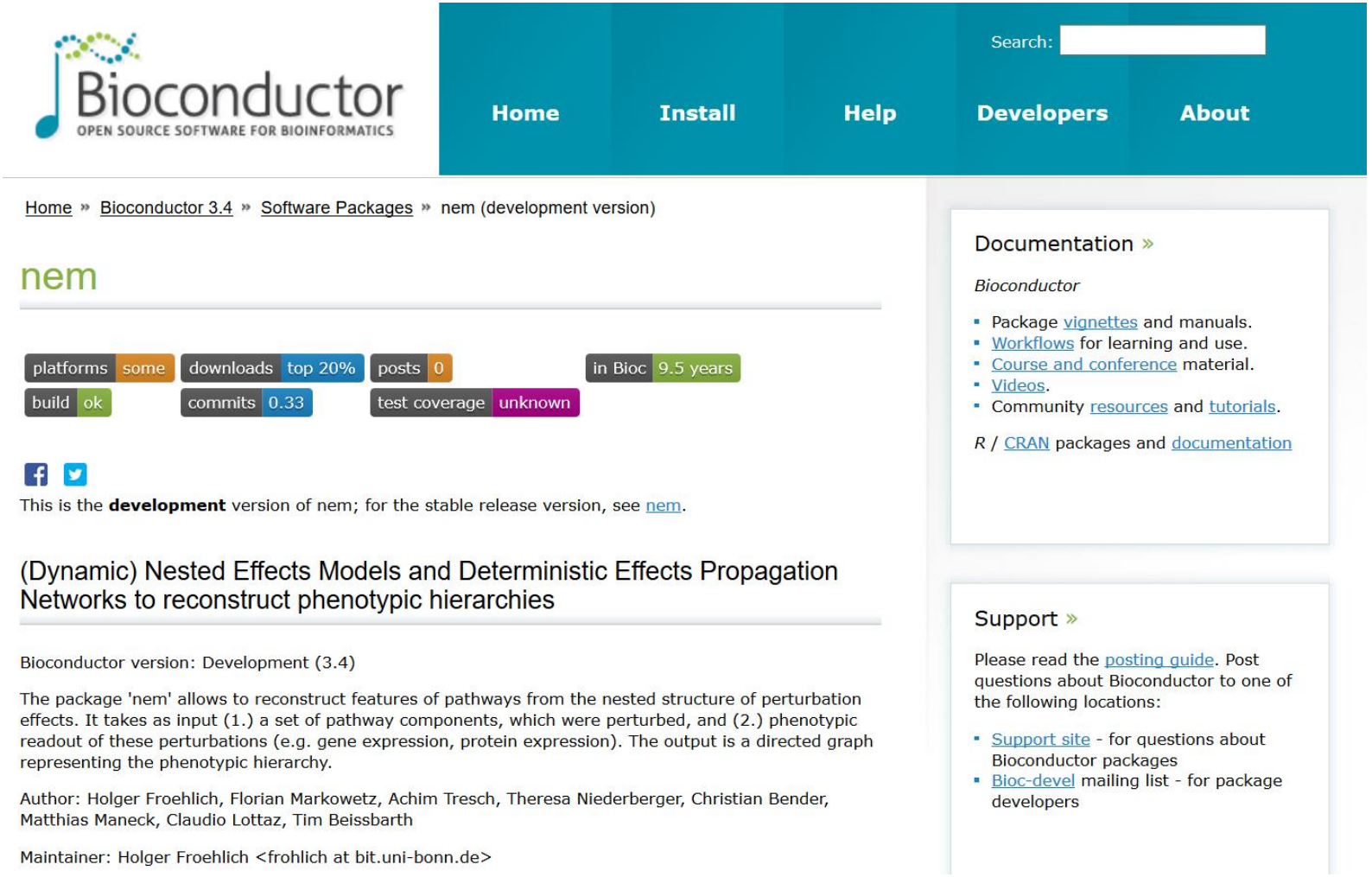
Associate Editor: Trey Ideker

ABSTRACT

Motivation: Targeted interventions in combination with the measurement of secondary effects can be used to computationally reverse engineer features of upstream non-transcriptional signaling cascades. *Nested effect models* (NEMs) have been introduced as a statistical approach to estimate the upstream signal flow from

change at gene *j*. Wagner (2001) uses such disruption networks as a starting point for a further graph-theoretic method, which removes indirect effects (Aho *et al.*, 1972), hence making the network more parsimonious. Tresch *et al.* (2007) and Klamt *et al.* (2010) enhance this approach by additionally making use of edge probabilities and signs to make the network consistent with the observed biological

Software for dNEM and dDEPN



Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

Search:

Home | Install | Help | Developers | About

Home » Bioconductor 3.4 » Software Packages » nem (development version)

nem

platforms some | downloads top 20% | posts 0 | in Bioc 9.5 years
build ok | commits 0.33 | test coverage unknown

This is the **development** version of nem; for the stable release version, see nem.

(Dynamic) Nested Effects Models and Deterministic Effects Propagation Networks to reconstruct phenotypic hierarchies

Bioconductor version: Development (3.4)

The package 'nem' allows to reconstruct features of pathways from the nested structure of perturbation effects. It takes as input (1.) a set of pathway components, which were perturbed, and (2.) phenotypic readout of these perturbations (e.g. gene expression, protein expression). The output is a directed graph representing the phenotypic hierarchy.

Author: Holger Froehlich, Florian Markowetz, Achim Tresch, Theresa Niederberger, Christian Bender, Matthias Maneck, Claudio Lottaz, Tim Beissbarth

Maintainer: Holger Froehlich <frohlich at bit.uni-bonn.de>

Documentation »

Bioconductor

- Package vignettes and manuals.
- Workflows for learning and use.
- Course and conference material.
- Videos.
- Community resources and tutorials.

R / CRAN packages and documentation

Support »

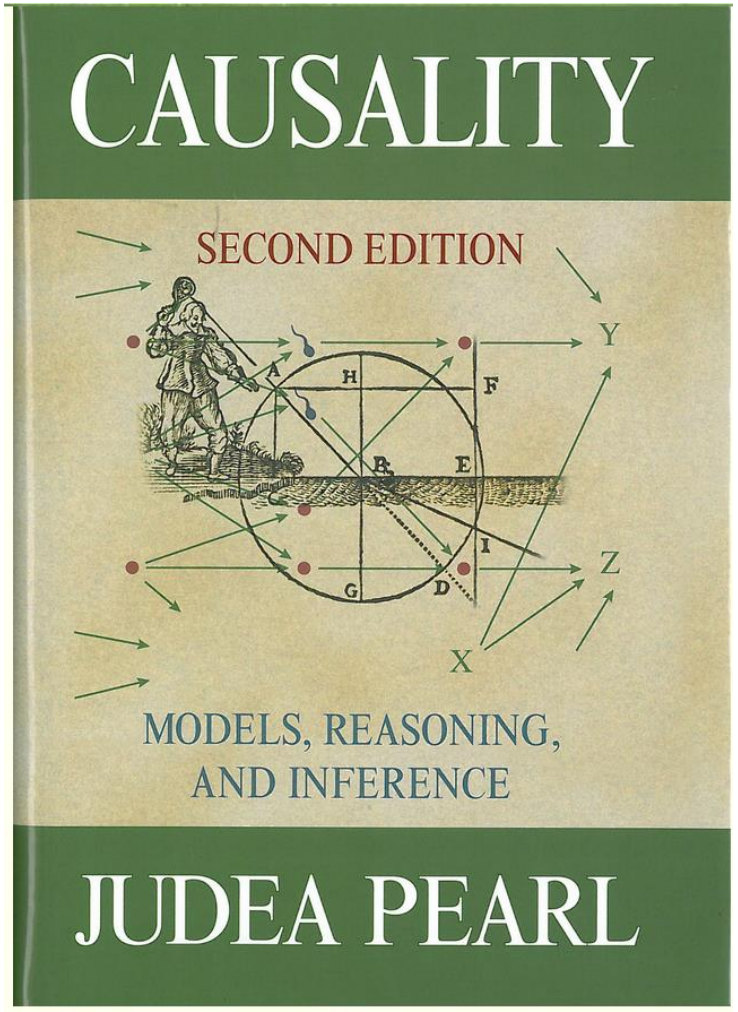
Please read the posting guide. Post questions about Bioconductor to one of the following locations:

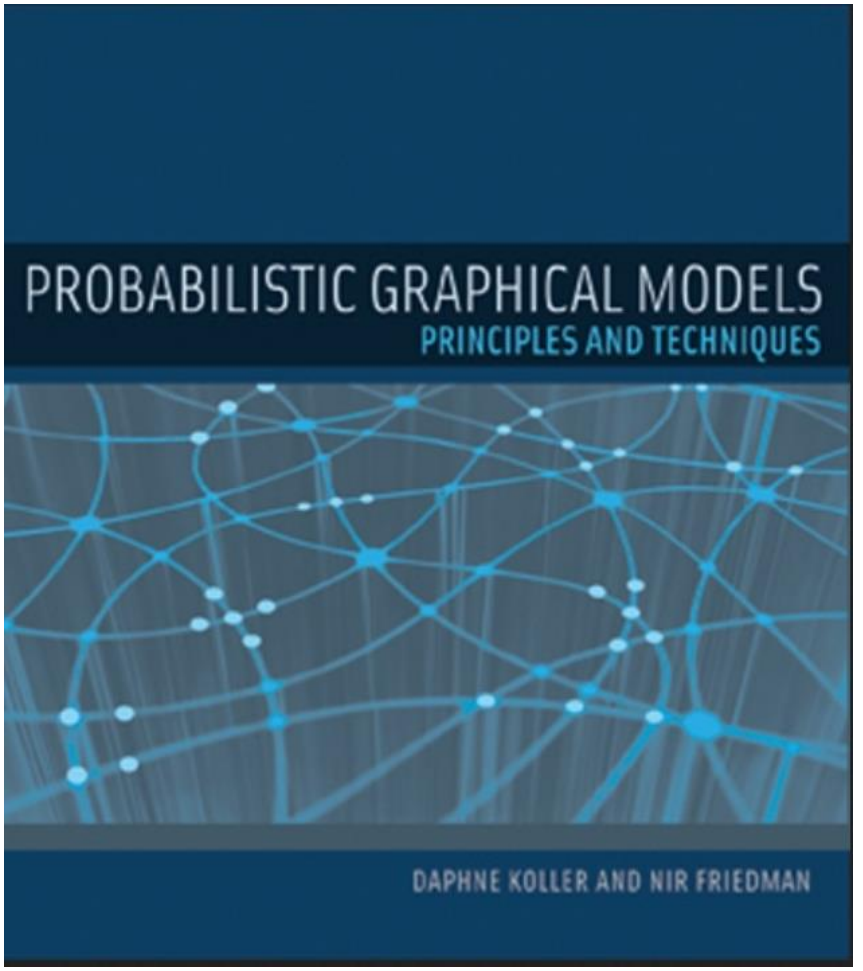
- Support site - for questions about Bioconductor packages
- Bioc-devel mailing list - for package developers

Bayesian Networks in Machine Learning (Past and Future)

- Backbone of Probailistic Machine Learning (since 90s)
- Most probabilistic algorithms in machine learning can be described in the language of Bayesian Networks
- Learning Bayesian networks (Belief networks) is one of the most important ML algorithms (along with Neural Networks and Gaussian Processes).
- Huge interest has led to seminial books written in this area.
- Deep Belief networks (*Hinton, G. (2009*)) extend Bayesian Networks towards Deep learning.

Bayesian Networks without tears





Thankyou