# BNfinder: free software for effective Bayesian Network inference

Bartek Wilczynski <wilczyns@embl.de>[*]

URL `http://launchpad.net/bnfinder`

code `http://code.launchpad.net/bnfinder/trunk`

license GNU General Public License

In recent years, we have seen an increased interest in applications of Bayesian Networks (BNs) in modelling in molecular biology. It is not surprising, since BNs are very natural models for reconstructing dependencies between observables, especially when measurements are noisy. They have been particularly successful in the field of regulatory genomics[1, 6], where the need for uncovering causal relationships between different variables is crucial.

While BNs are very flexible and naturally interpretable models, they are inherently difficult to train[2]. So far, the most common approach to training Bayesian networks were methods based either on Monte Carlo Markov chains[5] or on Expectation Maximization[6]. These approaches share common problems – they only provide approximate solutions while being computationally intensive. Recently, it has been shown by Dojer[3] that, under additional assumptions, the optimal solution to BN reconstruction problem can be found in polynomial time.

BNfinder[7] is a python implementation of this exact and efficient algorithm for BN reconstruction. It is distributed under GNU General Public License, so it can be freely used or adapted by other researchers. Since our project is currently still in its early stages, we are very open to any comments, suggestions or code contributions coming from potential users.

In this talk, I will briefly describe current state of the project and its design. Then I will describe three different examples of real problems that can be solved with BNfinder:

- reconstructing small gene networks from expression measurements both wildtype[8] and under perturbation[4]

- finding informative sequence motifs for groups of coexpressed genes[1]

- predicting gene expression from cis-regulatory modules (in preparation)

# References

[1] Michael A Beer and Saeed Tavazoie. Predicting gene expression from sequence. *Cell*, 117(2):185–198, Apr 2004.

[2] David Maxwell Chickering. Learning bayesian networks is np-complete. In *Learning from Data: Artificial Intelligence and Statistics V*, pages 121–130. Springer-Verlag, 1996.

[3] Norbert Dojer. Learning Bayesian networks does not have to be NP-hard. *LNCS*, 4162:305–314, 2006.

[4] Norbert Dojer, Anna Gambin, Bartek Wilczyński, and Jerzy Tiuryn. Applying dynamic Bayesian networks to perturbed gene expression data. *BMC Bioinformatics*, 7:249, 2006.

[5] Dirk Husmeier. Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian networks. *Bioinformatics*, 19(17):2271–82, 2003.

[6] E. Segal, M. Shapira, A. Regev, D. Pe'er, D. Botstein, D. Koller, and N. Friedman. Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat Genet*, 34(2):166–76, 2003.

[7] Bartek Wilczynski and Norbert Dojer. Bnfinder: Exact and efficient method for learning bayesian networks. *Bioinformatics*, page btn505, September 2008.

[8] Bartek Wilczyński and Jerzy Tiuryn. Reconstruction of mammalian cell cycle regulatory network from microarray data using stochastic logical networks. *Lecture Notes in Bioinformatics*, 4695/2007:121–135, 2007.

[*]EMBL Heidelberg and Institute of Informatics, University of Warsaw