

# Bayesian blind source separation of microarray data using graph-delayed correlation

Katrin Illner<sup>1</sup>, Christiane Dargatz<sup>1</sup>, Fabian J. Theis<sup>1</sup>

## Abstract

In the field of microarray analysis we often deal with heterogeneous data like mRNA, miRNA and protein level and in most cases we already know some interactions and dependencies between these variables. Many methods from the context of blind source separation, where we aim to find underlying biological processes (sources) and assume that the measurements (observations) are linear mixtures of these sources, do not consider any dependency structure. In the proposed work we therefore focus on this aspect and explicitly include the dependency of variables as prior knowledge in our model.

The idea comes from the context of time-series data and we deal with two main assumptions, first that the process is wide sense stationary, i.e. the time-delayed covariance is independent of the current time point, and second, that different sources have vanishing time-delayed covariance, cf. [3]. We transfer this concept to the framework of networks and introduce a graph-delayed covariance. Since we assume this covariance to be independent of the considered variable, we are able to estimate the strength of the dependency between variables in our model.

A first realization of the idea of a graph-delayed covariance was the GraDe-algorithm [2] as a sample-based approach. In the proposed work we now develop a probabilistic framework. This leads to several advantages, e.g. we can easily include prior information for the parameters and furthermore we can apply the concept of automatic relevance determination to decide which sources are the interesting and important ones. In more detail the method consists of a maximum a posteriori approach and we formulate the expectation maximization algorithm for our model to get estimates of the sources in a natural way. We calculate explicit update formulas for the single steps in the EM algorithm to ensure a fast computation. For first simulations and results we assume a line graph as dependency structure and show the application on FRAP-data.

---

<sup>1</sup>Institute for Bioinformatics and Systems Biology, Helmholtz Zentrum Munchen, Neuherberg, Germany

## References

- [1] C.M. Bishop. Bayesian pca. In *Advances in neural information processing systems 11: proceedings of the 1998 conference*, page 382. The MIT Press, 1999.
- [2] A. Kowarsch, F. Blöchl, S. Bohl, M. Saile, N. Gretz, U. Klingmuller, and F.J. Theis. Knowledge-based matrix factorization temporally resolves the cellular responses to IL-6 stimulation. *BMC bioinformatics*, 11(1):585, 2010.
- [3] L. Tong, VC Soon, YF Huang, and R. Liu. Amuse: a new blind identification algorithm. pages 1784–1787, 1990.