Contributed session 3

Ridge estimation of the VAR(1) model and its time series chain graph from multivariate time-course gene expression data

Viktorian Miok, Saskia M. Wilting, Wessel N. van Wieringen

To understand the crucial changes in the transcriptome during cervical carcinogenesis, we investigate temporal and contemporaneous gene-gene interactions. Cervical carcinogenesis can be faithfully mimicked morphologically and (epi)genetically in vitro . Our in vitro model system comprises four cell lines, two aected with HPV16 and two with HPV18. The cell lines are followed over time. At eight time points a sample of each cell line is proled for mRNA gene expression using microarrays.
The time-course gene expression data are described by a rst-order vector auto-regressive (VAR(1)) model. The auto-regressive and precision parameters of the VAR(1) model represent temporal and contemporaneous interaction among the genes within the time-series chain graph.
The parameters of the VAR(1) model are estimated by ridge penalized maximum likelihood. Optimal penalty parameters are obtained by cross-validation. Support of the time-series chain graph is determined by a post-estimation testing procedure. With knowledge of non-existent interactions among the genes, less biased estimates of the strengths of temporal and contemporaneous edges of the time-series chain graph are obtained using a constrained version of the ridge ML estimator. Several down-stream analyses for exploiting the reconstructed time-series chain graph are presented: node statistics, impulse response analysis, mutual information analysis, path decomposition and motifs illustration.
In a simulation study the proposed ridge ML estimator of the VAR(1) model outperforms a sparse competitor in terms of Frobenius loss of the estimates. With respect to the sensitivity and specicity of the edge selection the methods are on par.
The method is illustrated using data from the afore mentioned cervical cancer experiment mapping to the p53 signaling pathway (from KEGG repository). Central genes in the time-series chain graph reconstructed from these data are well-known drivers of the pathway associated with HPV induced transformation.
An implementation in R is forthcoming.

An empirical Bayes approach to network recovery using external knowledge

Gino Bertrand Kpogbezan

Networks reconstuction may benefit from inclusion of prior topological knowledge. Previously we employed a Bayesian Structural Equation Model (BSEM) for network reconstruction. In combination with variational Bayes it is flexible and computationally fast, and outperforms known competitors like glasso. Particular appealing of our BSEM approach is an empirical Bayes procedure for hyperparameter estimation.
In many settings external information on the topology of the network is available. For instance, from pathway repositories like KEGG in case of gene interaction networks, or inferred from data of a pilot study. Here we extend our BSEM to take such external data into account. The benefit of our approach are illustrated by analyzing gene expression data from GEO (Gene Expression Omnibus). We showed that our network estimates are clearly more reliable than those of competing methods by studying reproducibility of edges.

Identification of marginal causal relationships in gene networks, from observational and interventional expression data

Gilles Monneret, Florence Jaffrézic, Andrea Rau, Grégory Nuel

In this work, we identify a novel, objective, and targeted approach to identify candidates of causal downstream relationships in gene expression experiments consisting of replicated observational and partially available knock-out data (e.g., a knock-out of a single gene G). In particular, our approach proceeds in two steps. First, a correlation test is performed to identify genes that potentially interact with G; note that this initial group of genes may include those that are casually upstream or downstream of G, as well as those with spurious

correlations. In a second step, in order to deconvolute these possible relationships, we define a novel causal test to identify the partial order for each of the interaction pairs (G with the genes identified in the previous step). The proposed procedure is very fast and can be applied to thousands of genes simultaneously, which allows the pre-selection of a group of genes of interest for downstream causal network inference around gene G. An R package is under development to make the proposed causal selection approach easily applicable.