Estimation of networks and graphical structures

Mikko Sillanpää

University of Oulu

In Gaussian graphical models, assuming a multivariate normal distribution with unknown covariance matrix is a starting point for network estimation. Undirected network is typically constructed from the estimated precision matrix (inverse of the covariance matrix) which elements exhibit conditional independences and dependences between variables. This is, non-diagonal elements of precision matrix are zero (and non-zero) at the same time than the corresponding partial correlation coefficient is zero (and non-zero), respectively. This means in practice something like if network is estimated based on the covariance matrix rather than precision matrix, one may obtain larger number of connections that are resulting from indirect associations than direct ones. However, this is not to say that only causative associations are included to the network constructed based on precision matrix.

In biological applications, number of variables (nodes) are usually much higher than number of samples, which necessitates use of different regularization methods, which may, e.g., assume sparsity - that many of the precision matrix elements will be zero. A small tutorial to the topic is provided and some of the Bayesian and frequentist methods are shortly mentioned and discussed.