Are reconstructed molecular networks reproducible?

Wessel N. van Wieringen[1] (joint work with Yao Chen).

[1]Department of Epidemiology & Data Science, Amsterdam UMC, Amsterdam, The Netherlands
[2]Department of Mathematics, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

E-mail for correspondence: `w.vanwieringen@amsterdamumc.nl`

Abstract: Starting from the observation that the reconstruction of the cohesion among the variates of a multivariate random variable by means of Gaussian graphical model usually takes only sampling variation into account, we point out the consequences of this practice for the reconstruction of the underlying conditional independence graph. When replicates are included in the study, these consequences are overcome by the separation of sampling from other sources of variation. Hereto a simple 'signal+noise' model for the description of the multivariate data has been put forward. We present a penalized EM algorithm for the estimation of the model's parameters, alongside a discussion of cross-validation for choosing the penalty parameter(s). Through simulation we investigate how much is won by the inclusion of replicates, and compare the presented method to obvious alternatives. Finally, in an illustration using oncogenomics studies with replicates, we further investigate the effect of ignoring variation due to other sources than sampling variation and assess the reproducibility of the reconstruction of the conditional independence graph.

Key words: Conditional Independence Graph; Inverse covariance; Network; Reproducibility; Ridge penalty.

van Wieringen WN, Chen, Y (2020). Penalized estimation of the Gaussian graphical model from data with replicates. *submitted*.