**Clusterpath Gaussian Graphical Modeling**

Daniel J.W. Touw\(^1\), Patrick J.F. Groenen\(^1\), Ines Wilms\(^2\), and Andreas Alfons\(^1\)

\(^1\) Erasmus School of Economics, Erasmus University Rotterdam, (e-mail: touw@ese.eur.nl, groenen@ese.eur.nl, alfons@ese.eur.nl)

\(^2\) Department of Quantitative Economics, Maastricht University, (e-mail: i.wilms@maastrichtuniversity.nl)

**Abstract:** Gaussian graphical models (GGMs) serve as a means of summarizing conditional dependencies among a set of \(p\) variables. Such models are structured as networks, in which nodes represent individual variables and edges denote the presence of conditional dependence between two variables. Estimating GGMs in cases where the sample size \(n\) is smaller than the number of variables \((n < p)\) can present a challenge. To address this issue, existing estimation methods frequently rely on applying regularization techniques to the edges within the network, with the aim of obtaining a sparse network where many variables are represented as conditionally independent (see, e.g., Cai et al., 2011; Friedman et al., 2008; Meinshausen & Bühlmann, 2006; Peng et al., 2009; Rothman et al., 2008; Yuan, 2010).

Nevertheless, relying solely on edge sparsity does have limitations. First, when the number of variables is substantially larger than the sample size \((n \ll p)\), the conditional dependencies between variables may become too weak to detect (Eisenach et al., 2020). Second, sparse GGMs that include many variables can still contain a substantial number of edges, making interpretation difficult (Grechkin et al., 2015). Last, real-world networks often exhibit more complex structures than mere edge sparsity (Heinävaara et al., 2016; Hosseini & Lee, 2016).

To overcome these challenges, node aggregation has emerged as a means to perform dimension reduction in GGMs (see, e.g., Hosseini & Lee, 2016; Pircalabelu & Claeskens, 2020; Tarzanagh & Michailidis, 2018; Wilms & Bien, 2022). For example, instead of estimating the conditional dependencies between all observed variables, one may be interested in identifying the dependencies among a smaller number of clusters that share the same behavior. To achieve this, we propose the clusterpath GGM (CGGM), a model-based convex clustering Gaussian graphical model that automatically clusters groups of variables by means of the penalty structure used in the convex clustering literature (Hocking et al., 2011; Lindsten et al., 2011; Pelckmans et al., 2005).

**Keywords:** convex clustering, dimension reduction, graphical modeling, regularization
References


2, 494–515.