



## ΚΥΚΛΟΣ ΣΕΜΙΝΑΡΙΩΝ ΣΤΑΤΙΣΤΙΚΗΣ ΦΕΒΡΟΥΑΡΙΟΣ 2020

**Maarten Marsman**

*Assistant Professor, Department of Psychological Methods, University of Amsterdam*

### **Bayesian Variable Selection for the Ising Network Model with Applications in Psychology**

ΠΕΜΠΤΗ 20/2/2020, 13:15

**ΑΙΘΟΥΣΑ Τ103, ΤΡΟΙΑΣ 2, ΝΕΟ ΚΤΙΡΙΟ ΟΠΑ**

#### **ΠΕΡΙΛΗΨΗ**

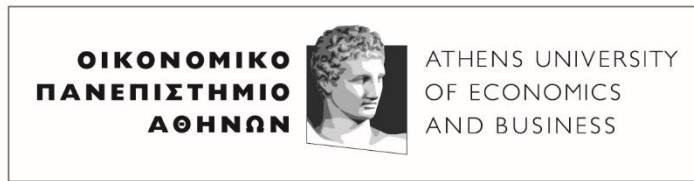
In the past few years, graphical models have become a popular tool in the psychological and educational sciences. In particular, the Ising model, a graphical model that originates from the study of magnetism in statistical physics, has received much attention in the psychometric literature. The Ising model describes the joint distribution of binary random variables in terms of main effects and pairwise associations. These binary variables could constitute the responses to items in an educational or intelligence test, or the symptoms of depression, and are the nodes of the network. The pairwise associations constitute the connections between the nodes in the network. Viewed in this way, a variable such as sleep loss can influence another variable such as a depressed mood if there is a connection between them. Therefore, a fundamental question in psychometric analyses is: Which variables influence each other, i.e., what is the network's structure?

Since the number of possible connections in the network grows quadratically with the number of variables, regularization is typically used to find parsimonious models. Based on different assumptions about the unknown network, two regularization methods can be distinguished: Low-rank methods assume that the network is densely connected, and Lasso-based methods assume that the network is sparsely connected. Whereas low-rank approaches have become more prevalent in intelligence and educational research, the Lasso approach has become more prevalent in psychopathology research.

Classical Lasso regularization is currently used in the analysis of relatively sparse network structures and combined with the extended BIC criterion to select (or delete) connections. This procedure comes with several problems. The exclusion of a connection with this procedure, for example, is often wrongfully taken for evidence that the connection at hand does not exist. Using Bayesian methods such evidence can

be expressed using, for example, posterior inclusion probabilities or Bayes factors. Another issue is the difficulty with computing standard errors with Lasso regularization. This, again, is natural in a Bayesian framework.

We develop a Bayesian framework for the analysis of sparse Ising networks. Bayesian variable selection methods are used to select the network's edges. An EM approach is developed to find interesting submodels, and a Gibbs sampling approach is developed to explore these submodels further. We show that our procedure is model selection consistent when the variance of the variable selection priors shrinks sufficiently fast. On top of this, our approach is simple, relatively fast, and easily accommodates missing data.



## AUEB STATISTICS SEMINAR SERIES FEBRUARY 2020

**Maarten Marsman**

*Assistant Professor, Department of Psychological Methods, University of Amsterdam*

### **Bayesian Variable Selection for the Ising Network Model with Applications in Psychology**

THURSDAY 20/2/2020, 13:15

**Room T103, 2 TROIAS STR., NEW AUEB BUILDING**

#### **ABSTRACT**

In the past few years, graphical models have become a popular tool in the psychological and educational sciences. In particular, the Ising model, a graphical model that originates from the study of magnetism in statistical physics, has received much attention in the psychometric literature. The Ising model describes the joint distribution of binary random variables in terms of main effects and pairwise associations. These binary variables could constitute the responses to items in an educational or intelligence test, or the symptoms of depression, and are the nodes of the network. The pairwise associations constitute the connections between the nodes in the network. Viewed in this way, a variable such as sleep loss can influence another variable such as a depressed mood if there is a connection between them. Therefore, a fundamental question in psychometric analyses is: Which variables influence each other, i.e., what is the network's structure?

Since the number of possible connections in the network grows quadratically with the number of variables, regularization is typically used to find parsimonious models. Based on different assumptions about the unknown network, two regularization methods can be distinguished: Low-rank methods assume that the network is densely connected, and Lasso-based methods assume that the network is sparsely connected. Whereas low-rank approaches have become more prevalent in intelligence and educational research, the Lasso approach has become more prevalent in psychopathology research.

Classical Lasso regularization is currently used in the analysis of relatively sparse network structures and combined with the extended BIC criterion to select (or delete) connections. This procedure comes with several problems. The exclusion of a connection with this procedure, for example, is often wrongfully taken for evidence that the connection at hand does not exist. Using Bayesian methods such evidence can be expressed using, for example, posterior inclusion probabilities or Bayes factors. Another issue is the difficulty with computing standard errors with Lasso regularization. This, again, is natural in a Bayesian framework.

We develop a Bayesian framework for the analysis of sparse Ising networks. Bayesian variable selection methods are used to select the network's edges. An EM approach is developed to find interesting submodels, and a Gibbs sampling approach is developed to explore these submodels further. We show that our procedure is model selection consistent when the variance of the variable selection priors shrinks sufficiently fast. On top of this, our approach is simple, relatively fast, and easily accommodates missing data.