Book of the Short Papers

Editors: Francesco Maria Chelli, Mariateresa Ciommi, Salvatore Ingrassia, Francesca Mariani, Maria Cristina Recchioni

SIS Società Italiana di Statistica

UNIVERSITÀ POLITECNICA DELLE MARCHE

Dipartimento di Scienze Economiche e Sociali DISES

LIUC Università Cattaneo

LIUC Università Cattaneo | BUSINESS ANALYTICS AND DATA SCIENCE HUB

Pearson

CHAIRS

Salvatore Ingrassia (Chair of the Program Committee) - *Università degli Studi di Catania*
Maria Cristina Recchioni (Chair of the Local Organizing Committee) - *Università Politecnica delle Marche*

PROGRAM COMMITTEE

Salvatore Ingrassia (Chair), Elena Ambrosetti, Antonio Balzanella, Matilde Bini, Annalisa Busetta, Fabio Centofanti, Francesco M. Chelli, Simone Di Zio, Sabrina Giordano, Rosaria Ignaccolo, Filomena Maggino, Stefania Mignani, Lucia Paci, Monica Palma, Emilia Rocco.

LOCAL ORGANIZING COMMITTEE

Maria Cristina Recchioni (Chair), Chiara Capogrossi, Mariateresa Ciommi, Barbara Ermini, Chiara Gigliarano, Riccardo Lucchetti, Francesca Mariani, Gloria Polinesi, Giuseppe Ricciardo Lamonica, Barbara Zagaglia.

ORGANIZERS OF INVITED SESSIONS

Pierfrancesco Alaimo Di Loro, Laura Anderlucci, Luigi Augugliaro, Ilaria Benedetti, Rossella Berni, Mario Bolzan, Silvia Cagnone, Michela Cameletti, Federico Camerlenghi, Gabriella Campolo, Christian Capezza, Carlo Cavicchia, Mariateresa Ciommi, Guido Consonni, Giuseppe Ricciardo Lamonica, Regina Liu, Daniela Marella, Francesca Mariani, Matteo Mazziotta, Stefano Mazzuco, Raya Muttarak, Livia Elisa Ortensi, Edoardo Otranto, Ilaria Prosdocimi, Pasquale Sarnacchiaro, Manuela Stranges, Claudia Tarantola, Isabella Sulis, Roberta Varriale, Rosanna Verde.

FURTHER PEPOPLE OF LOCAL ORGANIZING COMMITTEE

Elisa D'Adamo, Christian Ferretti, Giada Gabbianelli, Elvina Merkaj, Luca Pedini, Alessandro Pionati, Marco Tedeschi, Francesco Valentini, Rostand Arland Yebetchou Tchounkeu

Technical support: Matteo Mercuri, Maila Ragni, Daniele Ripanti

PUBLISHED BY PEARSON
WWW.PEARSON.COM
ISBN 9788891935618AAVV

Contents

Statistical Machine Learning for environmental applications

Statistical Process Monitoring for Complex Data in Industry 4.0

Advances in Data Science and Statistical Learning [IMS Invited Session]

ENBIS Session: System Maintenance, Boosting algorithms for regression, and Research Excellence

Educational Data mining: methods for complex data in students' assessment

Spatial and Spatio-Temporal Modeling: Theory and Applications

Statistical Framework for Measuring the Sustainability of Tourism

Statistical learning for well-being analysis

3 Contributed Sessions 305

Bayesian nonparametric methods

Economics and Statistics

Health statistics 1

Statistics in Society 1

Assessment and Education

Mixture Models

Sampling methods and analysis of survey data

Space-time statistics

Clustering and classification 1

Dynamic models and time series

Environmental learning and indicators

Health statistics 2

Migrants in Italy and return migration

Sustainability assessment

Bayesian methods and applications 2

Clustering and classification 2

Economics and labour markets

Environmental modeling

Multivariate data analysis 2

Statistics in Society 2

Bayesian methods and applications 3

Functional Data Analysis

Machine Learning and text mining

Multivariate data analysis 3

Quantile-based graphical models for continuous and discrete variables

Luca Merlo[a], Marco Geraci[b], and Lea Petrella[b]

[a]Department of Human Sciences, European University of Rome, Italy; luca.merlo@unier.it
[b]MEMOTEF Department, Sapienza University of Rome, Italy; marco.geraci@uniroma1.it, lea.petrella@uniroma1.it

Abstract

In this paper we develop a mixed graphical model for identifying conditional independence relations between continuous and discrete variables in a quantile framework using Parzen's definition of mid-quantile. To recover the graph structure and induce sparsity, we consider the neighborhood selection approach in which conditional mid-quantiles of each variable in the network are modeled as a sparse function of all others. Building on previous work, we propose a two-step estimation procedure where, in the first step, conditional mid-probabilities are obtained and, in the second step, the model parameters are estimated by solving an implicit equation with a LASSO penalty. The empirical application investigates the relationship between depression and inflammation on a sample of individuals from the National Health and Nutrition Examination Survey 2017-2020.

Keywords: LASSO, mixed random variables, mid-CDF, neighborhood selection, NHANES

1. Introduction

Graphical models have become a popular and effective framework for the statistical analysis of complex dependence relations among variables. Within this literature, Gaussian Graphical Models (GGMs, 11) have received considerable attention as they provide a model for the pair-wise conditional correlation structure of the variables of interest. Under the assumption of normality, the underlying conditional dependence structure is completely characterized by the inverse of the covariance matrix of the corresponding GGM. However, GGMs suffer from two limitations. First, they rely on the assumption of normally distributed data. Despite its simplicity and mathematical tractability, this assumption is hardly met in actual applications, and deviation from normality makes it harder to characterize conditional dependence structures. To overcome this issue, semi-parametric Gaussian copula models (15; 22) or power transformations of the data may be considered. Alternatively, one may forgo the normal distribution and consider more robust alternatives such as the multivariate t-distribution (7). Such proposals, however, have mainly relied on the use of symmetric distributions or on, more in general, location-shift models. In contrast, a quantile-based approach (1; 5) allows one to infer the conditional dependence structure without having to introduce assumptions on the form of the distributions.

Another, drawback of the GGMs is that they are confined to the modeling of continuous variables only. In many applications of practical relevance, however, the dataset of interest consists of mixed variables (categorical, counts, and continuous). Unfortunately, the literature regarding graphical models in which the variables are of different types is fairly limited. In parallel efforts, (23) and (3) introduced the class of Mixed Graphical Models (MGMs), which specify the conditional distribution of each variable (continuous and discrete) given the rest as a member of the exponential family of distributions. Subsequently, in a related line of research, (14) and (4) proposed a generalization of the conditional Gaussian model of (12) for mixed data.

The aim of this paper is to introduce a quantile-based graphical model for mixed variables that tackles conditional dependency structures, without making assumptions on the functional form of the distributions. We start from the work of (8) who developed a quantile regression method for discrete responses by extending Parzen's definition of marginal mid-quantiles (19). Intuitively, mid-quantiles can be viewed as fractional order statistics and have been extensively studied by (16). In this context, using mid-quantiles comes with desirable advantages as opposed to existing approaches, based on either jittering or latent constructs. Most importantly, they offer a unifying theory for quantile estimation with discrete or continuous variables, and are well-behaved asymptotically.

In our approach, to identify conditional independence relations and induce sparsity in the network, we model the conditional mid-quantiles of each variable as a sparse function of all others and fit separate regularized regressions using the neighborhood selection method of (17). For each variable, the parameters are estimated via a two-step procedure where conditional mid-probabilities are first obtained semi-parametrically and then regression coefficients are estimated by solving a LASSO-penalized implicit equation. The proposed method allows us to embed in a common graphical framework both continuous (possibly, e.g., heavy-tailed, skewed, multimodals) and discrete (e.g., binary, ordinal, count) variables, thus offering a much richer class of conditional distribution estimates than the conditional mean.

The relevance of this methodology is shown using observations from adult participants of the National Health and Nutrition Examination Survey (NHANES) 2017-2020 to investigate the association between C-Reactive Protein (CRP) and depression symptoms.

The rest of this paper is organized as follows. Sect. 2. formally describes the proposed model while the estimation procedure is discussed in Sect. 3. Finally, the empirical application is presented in Sect. 4.

2. Methods

In this section we illustrate the proposed mid-quantile mixed graphical model. In order to introduce our methods, we extend the mid-quantile regression of (8) to the graphical modeling framework with both continuous and discrete variables. Subsequently, using the neighborhood selection approach of (17), we show how to estimate a sparse mixed graphical model characterizing conditional independence relations among variables via node-wise penalized mid-quantile regressions.

Let $\mathbf{Y} = (X_1, \dots, X_{p_1}, Z_1, \dots, Z_{p_2})'$ denote a p-dimensional random vector, where $X_1, \dots, X_{p_1}$ are p_1 continuous variables and $Z_1, \dots, Z_{p_2}$ are p_2 discrete variables. Also, let $\mathcal{G} = (V, E)$ denote an undirected graph where $V = \{1, \dots, p\}$ is the set of nodes such that each component of the random variable $\mathbf{Y}$ corresponds to a node in V, and $E \subseteq V \times V$ represents the set of undirected edges.

Following (8), we first introduce the conditional mid-cumulative distribution function (mid-CDF, 19; 20) of Y_j given all other variables as

$$G_{Y_j|\mathbf{Y}_{\neg j}}(y_j \mid \mathbf{y}_{\neg j}) = F_{Y_j|\mathbf{Y}_{\neg j}}(y_j \mid \mathbf{y}_{\neg j}) - 0.5 m_{Y_j|\mathbf{Y}_{\neg j}}(y_j \mid \mathbf{y}_{\neg j}), \tag{1}$$

where $\mathbf{Y}_{\neg j}$ denotes all variables except Y_j, $F_{Y_j|\mathbf{Y}_{\neg j}}(\cdot \mid \cdot)$ is the conditional CDF of Y_j and $m_{Y_j|\mathbf{Y}_{\neg j}}(y_j \mid \mathbf{y}_{\neg j}) = \Pr(Y_j = y_j \mid \mathbf{Y}_{\neg j} = \mathbf{y}_{\neg j})$. The definition of conditional mid-CDF in eq. (1) applies to both continuous and discrete variables. Indeed, if Y_j is discrete, $G_{Y_j|\mathbf{Y}_{\neg j}}(y_j \mid \mathbf{y}_{\neg j})$ is a step function while it reduces to $F_{Y_j|\mathbf{Y}_{\neg j}}(y_j \mid \mathbf{y}_{\neg j})$ if Y_j is continuous since $\Pr(Y_j = y_j \mid \mathbf{Y}_{\neg j} = \mathbf{y}_{\neg j}) = 0$.

Let $\mathcal{S}_{Y_j}$ be the set of s distinct values in the population that the random variable Y_j can take on. Then, the conditional mid-quantile function (mid-QF) of Y_j, $H_{Y_j|\mathbf{Y}_{\neg j}}(\tau)$, is defined as the piecewise linear function connecting the values $G^{-1}_{Y_j|\mathbf{Y}_{\neg j}}(\pi_{jh} \mid \mathbf{y}_{\neg j})$, where $\pi_{jh} = G_{Y_j|\mathbf{Y}_{\neg j}}(y_j \mid \mathbf{y}_{\neg j})$, $h = 1, \dots, s$, for a given quantile level $\tau \in (0, 1)$. We model the τ-th conditional mid-quantile of Y_j given all the other variables using the following mid-quantile regression model:

$$H_{g_j(Y_j)|\mathbf{Y}_{\neg j}}(\tau) = \beta^0_j(\tau) + \mathbf{y}'_{\neg j}\boldsymbol{\beta}_j(\tau), \quad j = 1, \dots, p, \tag{2}$$

where $g_j(\cdot)$ is a known monotone and differentiable "link" function, and $\boldsymbol{\beta}_j(\tau) = (\beta_j^1(\tau), \ldots, \beta_j^{p-1}(\tau))'$ is a vector of $p-1$ unknown regression coefficients, with $\beta_j^0(\tau)$ being an intercept term, for a given τ.

To study conditional independence relations between the components of $\mathbf{Y}$ through the graph $\mathcal{G}$, we establish a result that allows us to make inference on the edge structure E using mid-quantile regressions. Following (1) and (5), the next proposition characterizes the relationship between the conditional mid-quantile function in eq. (2) and the conditional independence between any pair of variables in $\mathbf{Y}$ given the rest.

Proposition 1. *Suppose that the conditional mid-QF of a random variable Y_j, for some $j = 1, \ldots, p$, is defined by the mid-quantile regression model in eq. (2). Then, Y_j is conditionally independent from Y_k, with $k = 1, \ldots, p$ and $k \neq j$, given all of the other variables if and only if $\beta_j^k(\tau) = 0$ for all $\tau \in (0, 1)$.*

The proof of Proposition 1 follows from the relationship between the conditional mid-quantile and CDF of each node given the others. Most importantly, from Proposition 1 follows that the zero elements of the vector $\beta_j^k(\tau)$ for all $\tau \in (0, 1)$ identify conditional independence relations between the components of $\mathbf{Y}$. Hence, the edge set E of the graph $\mathcal{G}$ is completely determined by the non-zero components in the regression vector $\boldsymbol{\beta}_j(\tau)$, that is, $(j, k) \in E$ if and only if $\beta_j^k(\tau) \neq 0$. Based on this result, we can build a mixed quantile graphical model to characterize conditional independence relationships between the elements of $\mathbf{Y}$ by inferring the sparsity pattern of $\boldsymbol{\beta}_j(\tau)$.

We exploit the neighborhood selection approach of (17) by running separate mid-quantile regressions of each component in $\mathbf{Y}$ on all the others. Specifically, let $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_L)$ be a grid of L ordered quantile levels with $\tau_l \in (0, 1)$, $l = 1, \ldots, L$. Large values of L allow us to investigate conditional independence more accurately, but they also increase the computational cost of estimating the model. To infer the graph structure, we consider the linear model in eq. (2) for the conditional mid-QF, $H_{g_j(Y_j)|\mathbf{Y}_{\neg j}}(\tau_l)$, over all variables $j = 1, \ldots, p$ and levels $l = 1, \ldots, L$. Consequently, the corresponding edge set E of conditional dependencies is defined as

$$E = \Big\{(j, k) : \max_{l=1,\ldots,L} \{\max\{\mid \beta_j^k(\tau_l) \mid, \mid \beta_k^j(\tau_l) \mid\}\} > 0, \quad \text{for} \quad 1 \leq j \neq k \leq p\Big\}. \tag{3}$$

In the next section, we describe a procedure to estimate the proposed graphical model $\mathcal{G}$ and induce sparsity in the regression coefficients.

3. Estimation

Consider a sample $\mathbf{Y}_i, i = 1, \ldots, n$, with corresponding observations $\mathbf{y}_i$. For each variable Y_j, $j = 1, \ldots, p$ and level τ_l, $l = 1, \ldots, L$, estimation of the model in eq. (2), and in turn, of the set E in eq. (3), proceeds in two steps.

Let z_{jh}, $h = 1, \ldots, k$, be the h-th distinct observation of Y_j that occurs in the sample, with $z_{jh} < z_{jh+1}$ for all $h = 1, \ldots, k-1$. In the first step we estimate the mid-CDF in eq. (1), $\widehat{G}_{Y_j|\mathbf{Y}_{\neg j}}(y_j \mid \mathbf{y}_{\neg j})$, where $\widehat{F}_{Y_j|\mathbf{Y}_{\neg j}}$ is obtained by fitting h separate logistic regressions, one for each value of z_{jh}, $h = 1, \ldots, k$. In the second step, we define $\widehat{G}^c_{Y_j|\mathbf{Y}_{\neg j}}(y_j \mid \mathbf{y}_{\neg j})$ as the function interpolating the points $(z_{jh}, \widehat{G}_{Y_j|\mathbf{Y}_{\neg j}}(z_{jh} \mid \mathbf{y}_{\neg j}))$, where the ordinates have been obtained in the first step. The goal now is to estimate $(\beta^0(\tau_l), \boldsymbol{\beta}_j(\tau_l))$ in eq. (2) by solving the implicit equation $\tau_l = \widehat{G}^c_{Y_j|\mathbf{Y}_{\neg j}}(\eta(\tau_l) \mid \mathbf{y}_{\neg j})$, where $\eta(\tau_l) = g_j^{-1}\{\beta^0(\tau_l) + \mathbf{y}'_{\neg j}\boldsymbol{\beta}_j(\tau_l)\}$, and, at the same time, capture the most relevant interconnections between the variables, which motivates us to use a sparse estimator that automatically shrink the elements of $\boldsymbol{\beta}_j(\tau_l)$. Following (8), we thus obtain an estimate of $\boldsymbol{\beta}_j(\tau_l)$, denoted $\widehat{\boldsymbol{\beta}}_j(\tau_l)$, by minimizing the following objective function

$$\arg\min_{\boldsymbol{\beta}} \frac{1}{n} \sum_{i=1}^{n} \Big(\tau_l - \widehat{G}^c_{Y_j|\mathbf{Y}_{\neg j}}(\eta_i \mid \mathbf{y}_{\neg j})\Big)^2 + \lambda \mid\mid \mathrm{diag}(\mathbf{w})\boldsymbol{\beta}_j(\tau_l) \mid\mid_1, \tag{4}$$

where we consider the linear interpolation function

$$\widehat{G}^{c}_{Y_j|\mathbf{Y}_{\neg j}}(\eta_i \mid \mathbf{y}_{\neg j}) = b_{h_i}(\eta_i - z_{jh_i}) + \widehat{\pi}_{jh_i}, \quad z_{jh_i} \leq \eta_i \leq z_{jh_i+1}, \tag{5}$$

with $b_{h_i} = \frac{\widehat{\pi}_{jh_i+1}-\widehat{\pi}_{jh_i}}{z_{jh_i+1}-z_{jh_i}}$ and $\widehat{\pi}_{jh_i} = \widehat{G}_{Y_j|\mathbf{Y}_{\neg j}}(z_{jh_i} \mid \mathbf{y}_{\neg j})$. The penalization in eq. (4) is given by a Lasso-type penalty on $\boldsymbol{\beta}_j(\tau_l)$ where we allow a different weight for each coefficient by using the vector $\mathbf{w}$ to avoid that variables of different types are on different scales, and where $\lambda \geq 0$ is the overall tuning parameter of the model. The parameter λ controls the strength of the penalization and determines the sparsity of the graph: a higher (lower) value is responsible for a lower (higher) number of edges; when $\lambda = 0$, $\widehat{\boldsymbol{\beta}}_j(\tau_l)$ reduces to the closed-form estimator in (8, see eq. 2.9). Finally, to infer the graph structure we solve the minimization problem in eq. (4) for all Y_j, $j = 1, \ldots, p$ and τ_l, $l = 1, \ldots, L$, and estimate the edge set E as follows:

$$\widehat{E} = \Big\{(j,k) : \max_{l=1,\ldots,L}\{\max\{|\ \widehat{\beta^{k}_{j}}(\tau_l)\ |, |\ \widehat{\beta^{j}_{k}}(\tau_l)\ |\}\} > 0, \quad \text{for} \quad 1 \leq j \neq k \leq p\Big\}. \tag{6}$$

To select the optimal value of the penalty parameter λ, we adopt the following BIC-type criteria (5):

$$\text{BIC}(\lambda) = \sum_{l=1}^{L}\sum_{j=1}^{p}\left[\ln\left(\sum_{i=1}^{n}\rho_\tau(y_{ij} - \beta^0_j(\tau_l) - \mathbf{y}'_{i\neg j}\boldsymbol{\beta}_j(\tau_l))\right) + \frac{\ln n \ln(p-1)}{2n}\nu_{jl}\right], \tag{7}$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$ is the quantile loss function (9), with $I(\cdot)$ being the indicator function, and ν_{jl} is the number of estimated non-zero components in $\widehat{\boldsymbol{\beta}}_j(\tau_l)$ for node j at quantile level τ_l. Specifically, we fit the model for a grid of candidate values of λ and then select the optimal tuning parameter as that corresponding to the lowest BIC value in eq. (7).

4. Application

To evaluate the performance of the proposed methods, we illustrate an application to depression symptoms and inflammatory proteins from the NHANES 2017-2020. There is mounting evidence that inflammatory proteins adversely affect functional ability, quality of life, and well-being of individuals (18; 13). Among these proteins, C-Reactive Protein (CRP) is arguably the most extensively studied inflammatory index in depression research. CRP is a protein synthesized by the liver during the acute phase of an inflammatory/infectious process in response to stimulation from other pro-inflammatory proteins (e.g., elevated cytokine levels (6)). Research suggests that the presence, size, and direction of the association between CRP level and depression vary. One possible explanation for this might be the heterogeneity in the population due to, e.g., age and race/ethnicity. Another explanation may be that CRP is also associated with numerous factors (confounders), such as socio-demographic variables and the overall health status of the individual. As potential reasons for these inconsistencies, (21) also pointed out differences in population settings (e.g., inpatient, outpatient, or community), depression assessment (e.g., sum-scores or diagnoses), or adjustment for important chronic conditions.

Graphical models represent the ideal tool to help disentangling the intricate dependencies between CRP, depression symptoms and individual characteristics. Following (8), before carrying out the analysis we remove the effect of NHANES oversampling and then restrict the dataset to individuals aged 20-70 years. The final sample size for analysis is $n = 3690$, composed of about 59.4% of white persons and 50.1% females. In this network, we include the concentration of CRP (mg/L), nine depression symptoms measured via the Patient Health Questionnaire-9 (PHQ-9, 10) scored on a 4-point Likert scale and 17 socio-demographic, clinical, and lifestyle variables collected among the survey participants, resulting in a total of $p = 17$ nodes. Specifically, the PHQ-9 is a nine-item self-report questionnaire that was administered to assess the frequency of nine major depression criteria listed in the Diagnostic and Statistical Manual of Mental Disorders (2). The questionnaire is a well-established, validated tool that evaluates how often individuals had been bothered by any of the nine items in the previous 2 weeks, on a scale ranging from 0 (“not at all”) to 3 (“nearly every day”).

We fit the proposed model using a grid of $L = 7$ quantile levels, $\boldsymbol{\tau} = (1/8, \ldots, 7/8)$, across an equispaced sequence of 100 values of the tuning parameter λ on the log scale from 0.001 to 5. Prior to fitting, continuous variables have been centered around zero and divided by their standard deviation. For continuous and count variables, we take $g(\cdot)$ to be, respectively, the identity and the logarithmic function, while we use the logistic mid-quantile model for binary nodes. Finally, the edge set $\widehat{E}$ is estimated as described in eq. (6). To reduce model uncertainty and improve reliability of the inferred interactions, we adopt a model averaging approach. Specifically, 500 bootstrap datasets are created by resampling from the original one. Then, we fit the proposed model on the 500 bootstrap re-samples and estimate the edge structure for each bootstrap dataset. Eventually, in the final network we retain only those edges that are present in at least 85% (18) of the learned graphs.

Fig. 1 provides a graphical representation of the estimated network, where the width of the edges is proportional to the absolute value of the strength of the interaction and the edge colors specify the sign of the corresponding interaction (green = positive, red = negative, grey = undefined). The colors of the nodes map to the three different domains, Inflammation Marker, Depression Criteria, and Covariates.

Results indicate that CRP is associated with greater changes in appetite, presenting a non-zero edge in 90% of bootstrap re-samples. CRP also shows noteworthy connections with other variables: it is positively associated with BMI and gender but negatively with recent smoking and alcohol consumption. Finally, smoking and gender are proximal to several symptom criteria including fatigue, appetite problems, psychomotor changes and thoughts of death, suggesting that there may be gender differences underlying these relationships.

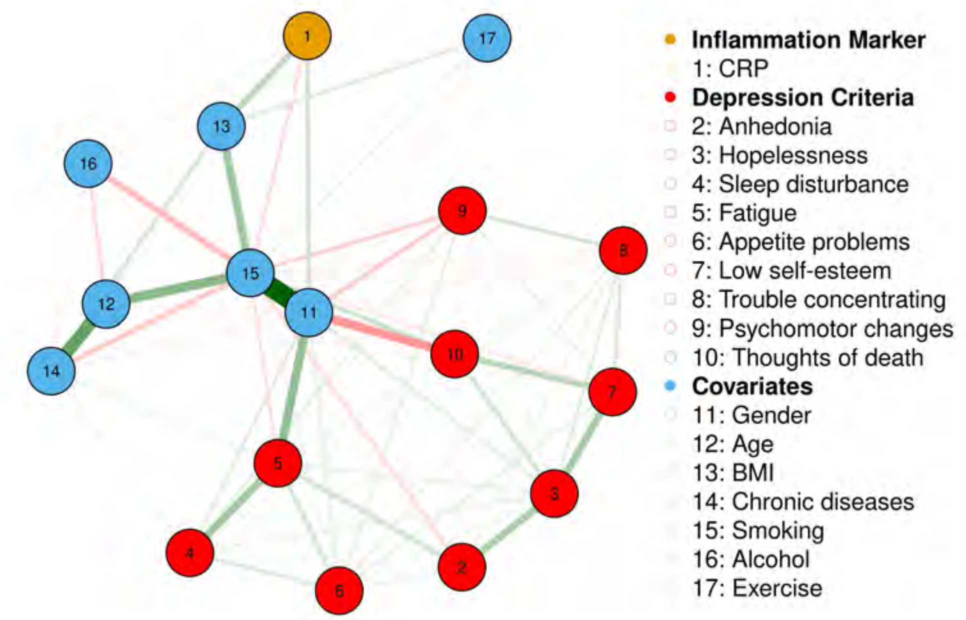

Figure 1: Estimated graph structure. Edges represent interactions that was found to be non-zero in at least 85% of the 500 bootstrap re-samples. Green edges in the networks depict positive associations, red edges represent negative associations, and grey ones signify interactions wherein no sign is defined. Thicker edges depict stronger associations.

References

[1] Ali, A., Kolter, J.Z., Tibshirani, R.J.: The multiple quantile graphical model. Adv. Neural. Inf. Process. Syst. **29** (2016)

[2] American Psychiatric Association, D., Association, A.P., et al.: Diagnostic and statistical manual of mental disorders: DSM-5, vol. 5. American Psychiatric Association Washington, DC (2013)

[3] Chen, S., Witten, D.M., Shojaie, A.: Selection and estimation for mixed graphical models. Biometrika **102**(1), 47–64 (2015)

[4] Cheng, J., Li, T., Levina, E., Zhu, J.: High-dimensional mixed graphical models. J. Comput. Graph. Stat. **26**(2), 367–378 (2017)

[5] Chun, H., Lee, M.H., Fleet, J.C., Oh, J.H.: Graphical models via joint quantile regression with component selection. J. Multivar. Anal. **152**, 162–171 (2016)

[6] Du Clos, T.W.: Function of C-reactive protein. Ann. Med. **32**(4), 274–278 (2000)

[7] Finegold, M., Drton, M.: Robust graphical modeling of gene networks using classical and alternative t-distributions. Ann. Appl. Stat. pp. 1057–1080 (2011)

[8] Geraci, M., Farcomeni, A.: Mid-quantile regression for discrete responses. Stat. Methods Med. Res. **31**(5), 821–838 (2022)

[9] Koenker, R., Bassett, G.: Regression Quantiles. Econometrica **46**(1), 33–50 (1978)

[10] Kroenke, K., Spitzer, R.L., Williams, J.B.: The PHQ-9: validity of a brief depression severity measure. J. Gen. Intern. Med. **16**(9), 606–613 (2001)

[11] Lauritzen, S.L.: Graphical models, vol. 17. Clarendon Press (1996)

[12] Lauritzen, S.L., Andersen, A.H., Edwards, D., Jöreskog, K.G., Johansen, S.: Mixed graphical association models. Scand. J. Stat. pp. 273–306 (1989)

[13] Lee, C., Min, S.H., Niitsu, K.: C-reactive protein and specific depression symptoms among older adults: An exploratory investigation of multi-plane networks using cross-sectional data from NHANES (2017–2020). Biol. Res. Nurs. **25**(1), 14–23 (2023)

[14] Lee, J.D., Hastie, T.J.: Learning the structure of mixed graphical models. J. Comput. Graph. Stat. **24**(1), 230–253 (2015)

[15] Liu, H., Han, F., Yuan, M., Lafferty, J., Wasserman, L.: High-dimensional semiparametric Gaussian copula graphical models. Ann. Stat. **40**(4), 2293–2326 (2012)

[16] Ma, Y., Genton, M.G., Parzen, E.: Asymptotic properties of sample quantiles of discrete distributions. Ann. Inst. Stat. Math. **63**(2), 227–243 (2011)

[17] Meinshausen, N., Bühlmann, P.: High-dimensional graphs and variable selection with the lasso. Ann. Stat. **34**(3), 1436–1462 (2006)

[18] Moriarity, D.P., Horn, S.R., Kautz, M.M., Haslbeck, J.M., Alloy, L.B.: How handling extreme C-reactive protein (CRP) values and regularization influences CRP and depression criteria associations in network analyses. Brain Behav. Immun. **91**, 393–403 (2021)

[19] Parzen, E.: Change PP plot and continous sample quantile function. Optim. **22**(12), 3287–3304 (1993)

[20] Parzen, E.: Quantile probability and statistical data modeling. Stat. Sci. pp. 652–662 (2004)

[21] Smith, K.J., Au, B., Ollis, L., Schmitz, N.: The association between C-reactive protein, Interleukin-6 and depression among older adults in the community: a systematic review and meta-analysis. Exp. Gerontol. **102**, 109–132 (2018)

[22] Xue, L., Zou, H.: Regularized rank-based estimation of high-dimensional nonparanormal graphical models. Ann. Stat. **40**(5), 2541–2571 (2012)

[23] Yang, E., Baker, Y., Ravikumar, P., Allen, G., Liu, Z.: Mixed graphical models via exponential families. In: Artif. Intell. and Stat., pp. 1042–1050. PMLR (2014)