Bayesian Regression with Network Predictors

Regressione Bayesiana con Covariate di Rete

Daniele Durante and David B. Dunson

Abstract Our focus is on prediction and inference on the association between a network-valued random variable and a response. The motivation is drawn from neuroscience studies measuring a brain connectivity network for each subject along with a response, such as an intelligence score. A recent nonparametric model for network data automatically clusters subjects into groups according to their brain network structure. We build on this model by proposing a Bayesian linear regression that allows the response conditional expectation to shift over the network clusters, facilitating inference on the association between the network and the response. A Gibbs sampler is defined for posterior computation. The approach is applied to data on human brain networks and intelligence scores.

Abstract *L'obiettivo di questo lavoro è quello di sviluppare modelli statistici per studiare relazioni tra una variabile risposta univariata ed un regressore di rete. Una particolare attenzione è prestata ad applicazioni nelle neuroscienze dove per ogni soggetto sono disponibili sia dati sulla sua rete cerebrale sia altre informazioni, tra cui un punteggio di intelligenza. Un recente modello nonparametrico consente di raggruppare reti in funzione di strutture di connessioni simili. Questo lavoro sfrutta la precedente metodologia proponendo un modello di regressione Bayesiana per la variabile risposta, il cui valore atteso condizionato varia attraverso i diversi cluster di rete. Viene proposto un Gibbs sampler per il calcolo delle distribuzioni a posteriori e il modello è applicato a dati su reti cerebrali umane e punteggi di intelligenza.*

Key words: Bayesian linear regression, Low-rank factorization, Mixture model, Network predictor, Neuroscience.

Daniele Durante
Department of Statistical Sciences, University of Padua, Padua, Italy. e-mail: durante@stat.unipd.it

David B. Dunson
Department of Statistical Science, Duke University, Durham, NC, USA. e-mail: dunson@duke.edu

1 Introduction

Networks represent interconnections among a set of nodes, and they are typically available via adjacency matrices, with each entry measuring a link among the corresponding pair of nodes. The typical focus in the literature on statistical methods for networks considers a single network observation; such models include ERGMs (Erdös and Rényi, 1959, Holland and Leinhardt, 1981, Frank and Strauss, 1986), stochastic block models (Nowicki and Snijders, 2001) and latent space models (Hoff et al., 2002). There has also been some focus on developing predictive models for network inference, with nodes or edge covariates informing on the interconnection structure (see e.g. Snijders et al., 2006, Hoff et al., 2002, Snijders, 2005, Durante and Dunson, 2014). These models study the conditional distribution of edges within a single network, exploiting covariate information. We instead aim to address a completely different problem in which there is a different network observed for each subject in a study, and we use the network as a predictor of subject-specific response variables.

Our contribution is specifically motivated by the attempt to understand whether variations in the brain network structure have an effect in explaining the intelligence process. Data are available via repeated measures of pairs $\{y_i, A_i\}$, $i = 1, \ldots, n$, with y_i an intelligence score for the ith individual measured via Full Scale Intelligence Quotient (FSIQ) and A_i the $V \times V$ symmetric adjacency matrix representing the corresponding brain network structure, so that $A_{i[vu]} = A_{i[uv]} = 1$ if a connection is observed among brain regions $v = 2, \ldots, V$ and $u = 1, \ldots, v-1$ in individual i and $A_{i[vu]} = A_{i[uv]} = 0$ otherwise; see Figure 1 for an illustrative example. There is an increasing emphasis within the neuroscience community towards the need to understand whether variations in the brain connectivity behavior are informative about phenotypic variability, and how such different network structures are related to higher or lower values of a given phenotype.

Previous proposals typically rely on descriptive analyses or modeling of specific network features, such as number of nodes, clustering coefficients and k-cores. In contrast, we provide a fully generative joint modeling approach, which (i) exploits network information without the need to reduce the interconnection structure to features prior to statistical analysis, (ii) allows network data to inform about the variability of a given response variable and (iii) infers which type of network structures are associated with lower or higher values of the response. The methodology relies on first clustering the network predictors via a recently developed Bayesian mixture of low-rank factorizations (Durante et al., 2015) so that networks in the same cluster share a similar connectivity behavior. The conditional distribution for the continuous response is then modeled via a Bayesian linear regression, with a specific intercept for each network cluster informing how specific brain network structures impact on the conditional expectation of the response.

Hence, our contribution is related to the literature on Bayesian joint modeling for functional predictors (Bigelow and Dunson, 2009). Bigelow and Dunson (2009) favor clustering among predictor trajectories, with each cluster associated to a specific offset in the generalized liner model (GLM) for the response variable. Although we

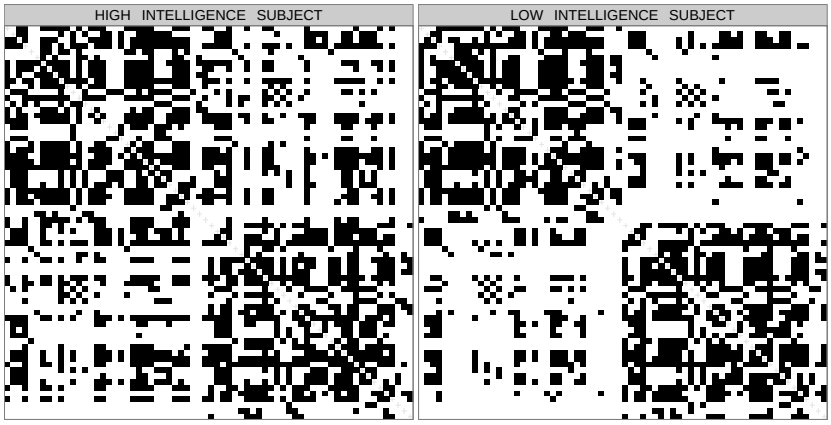

Fig. 1 Adjacency matrices representing selected brain networks. Black refers to an edge; white to a non-edge.

are similar to this contribution in developing a predictive model, which infers the dependence between a scalar response and a non-scalar predictor, there is a substantial difference in dealing with functional and network predictors, for which literature is still lacking.

The paper is organized as follows. In Section 2, we describe the model formulation and posterior computation via Gibbs sampling. Section 3 examines an application to understand the relation between brain network and intelligence.

2 Joint Modeling of Networks and Continuous Response

Letting $\{y_i, A_i\}$ denote the pair encoding a response variable and a network predictor, respectively, observed on individual i with $i = 1, \ldots, n$, we look for a simple and interpretable model, which allows the network A_i to be predictive of the response variable y_i. Considering our neuroscience motivating application it is appealing, for the sake of inference and interpretability, to group networks into clusters, with networks in the same group sharing the same connectivity behavior, and assess how the conditional expectation of the phenotypic intelligence response varies across such clusters.

2.1 Clustering Network Predictors via Mixture Models

Consistently with previous aims, we induce clustering among the network predictors A_i, $i = 1, \ldots, n$ exploiting a recently proposed mixture of low-rank factorizations

for network-valued random variables (Durante et al., 2015). Specifically, under the assumption of undirected network observations with no self-relationships as in our application, and letting $\mathscr{L}(A_i)$ the vector of the lower triangular elements of A_i, with $\mathscr{L}(A_i) = (A_{i[21]}, A_{i[31]}, \ldots, A_{i[V1]}, A_{i[32]}, \ldots, A_{i[V2]}, \ldots, A_{i[V(V-1)]})^T \in \{0,1\}^{V(V-1)/2}$, the model assumes the edges $\mathscr{L}(A_i)_l$, $l = 1, \ldots, V(V-1)/2$ as conditionally independent Bernoulli random variables given a latent cluster membership G_i with $\Pr(G_i = h) = \nu_h$, $h = 1, \ldots, H$ and a cluster-specific edge probability vector $\pi^{(h)} \in (0,1)^{V(V-1)/2}$, having entries $\pi_l^{(h)} = \Pr\{\mathscr{L}(A_i)_l = 1 \mid G_i = h\}$ for $l = 1, \ldots, V(V-1)/2$, $h = 1, \ldots, H$. Hence

$$\{\mathscr{L}(A_i)_l \mid \pi_l^{(h)}, G_i = h\} \sim \text{Bern}(\pi_l^{(h)}), \tag{1}$$

independently for $l = 1, \ldots, V(V-1)/2$ and $i = 1, \ldots, n$. To further reduce the dimensionality of the problem and favor borrowing of information within the network, each cluster-specific edge probability vector $\pi^{(h)}$, $h = 1, \ldots, H$ is defined via a latent space representation. In particular, each edge is assigned a shared propensity to be observed across all subjects networks as well as a deviation from this shared behavior which is common to all networks sharing the same cluster membership h. As there are less information in the data about cluster-specific deviations, the latter is constructed via matrix factorization procedures efficiently exploiting network information; see Durante et al. (2015) for specific details on the mixture of low rank factorization model and related priors for the corresponding quantities.

2.2 Dependent Response Model

Consistently with our aim of assessing how the phenotypic intelligence response depends on brain networks, we consider a Bayesian linear regression model for data $y_1, \ldots, y_n$, which includes a specific offset $\mu_h \in \Re$ for each possible network group h, $h = 1, \ldots, H$ arising from the previous clustering. Hence, we obtain

$$y_i \mid \mu_i, \sigma^2 \sim \text{N}(\mu_i, \sigma^2), \quad \text{with} \quad \mu_i = \sum_{h=1}^{H} \mu_h 1_{(h)}(G_i) + x_i^T \beta, \tag{2}$$

where $x_i = (x_{i1}, \ldots, x_{ip})^T$ is a vector of additional predictors of control, such as age or gender, $\beta = (\beta_1, \ldots, \beta_p)^T \in \Re^p$ are their corresponding coefficients and $1_{(h)}(G_i)$ is an indicator function such that $1_{(h)}(G_i) = 1$ if $G_i = h$ and $1_{(h)}(G_i) = 0$ otherwise.

Under model (2) different subpopulations of networks from the clustering in (1) are allowed to change the location of the expected value for the response. As a result, by jointly analyzing the posterior distributions for the cluster-specific edge probability vectors $\pi^{(h)}$ in (1) along with the cluster-specific intercepts μ_h, $h = 1, \ldots, H$ in the response model (2) we obtain simple inference strategies to learn if and which type of network structures relates to higher or lower values of a phenotypic response.

To conclude Bayesian specification we consider standard settings from Bayesian linear regression assuming priors $\theta = (\mu_1, \ldots, \mu_H, \beta_1, \ldots, \beta_p)^T \sim \mathrm{N}_{H+p}(\mu_\theta, \Sigma_\theta)$ and $\sigma^{-2} \sim \mathrm{Ga}(a,b)$. These choices allow the definition of a simple and efficient Gibbs sampler. The first step of the proposed MCMC consists in updating the cluster allocation for each subject G_i, $i = 1, \ldots, n$ from their full conditional discrete random variable with $\Pr(G_i = h \mid -)$ equal to

$$\frac{\left[\nu_h \prod_{l=1}^{V(V-1)/2} \{\pi_l^{(h)}\}^{\mathscr{L}(A_i)_l} \{1-\pi_l^{(h)}\}^{1-\mathscr{L}(A_i)_l}\right] \mathrm{N}(y_i; \mu_h + x_i^T \beta, \sigma^2)}{\sum_{m=1}^{H} \left[\nu_m \prod_{l=1}^{V(V-1)/2} \{\pi_l^{(m)}\}^{\mathscr{L}(A_i)_l} \{1-\pi_l^{(m)}\}^{1-\mathscr{L}(A_i)_l}\right] \mathrm{N}(y_i; \mu_m + x_i^T \beta, \sigma^2)},$$

for each $h = 1, \ldots, H$ and $i = 1, \ldots, n$. Given the cluster memberships G_i, $i = 1, \ldots, n$ the Gibbs sampler proceeds as in Durante et al. (2015) to update the cluster-specific edge probability vectors $\pi^{(h)}$, $h = 1, \ldots, H$ and consider standard results from Bayesian linear regression to sample from the full conditionals of θ and σ^{-2}.

The model outlined along with the proposed MCMC algorithm shows good estimation and mixing performance in several simulation settings. It is also important to notice that when trace-plots suggest label switching issues are encountered, it is necessary relabel the classes at each MCMC iteration using post-processing algorithms, such as Stephens (2000).

3 Brain Networks and FSIQ

We illustrate the performance of our model in equations (1)–(2) by considering an application to learn possible relations between brain networks A_i and intelligence scores y_i on a sample of $n = 113$ subjects.

Intelligence scores are available via Full Scale Intelligence Quotient (FSIQ), which provides an overall measure of cognitive ability summarizing perceptual reasoning, processing speed, verbal comprehension and working memory; see e.g. Koriakin et al. (2013) and the references cited therein. Brain networks are constructed by processing Diffusion Tensor Imaging (DTI) scans via recently developed pipelines (Roncal et al., 2013) to obtain symmetric binary adjacency matrices A_i having elements $A_{i[vu]} = A_{i[uv]} = 1$ if at least one white matter fiber connects brain regions $v = 2, \ldots, 68$ and $u = 1, \ldots, v-1$ in individual i and $A_{i[vu]} = A_{i[uv]} = 0$ otherwise. We have a total of 68 brain regions as defined by the Desikan Atlas (Desikan et al., 2006). For each subject, additional information on age and gender is available. We include these predictors in (2) as control variables.

Posterior computation is performed utilizing the Durante et al. (2015) hyperparameter settings for priors in the mixture model in (1). For Bayesian regression in (2), we consider instead diffuse priors with the group specific offsets centered on the sample mean of the intelligence response. Specifically we let $\mu_\theta = (\bar{y}, \ldots, \bar{y}, 0, 0)$, $\Sigma_\theta = \mathrm{diag}(100, \ldots, 100)$ and $a = b = 0.01$. Finally, we set $H = 3$ consistently with

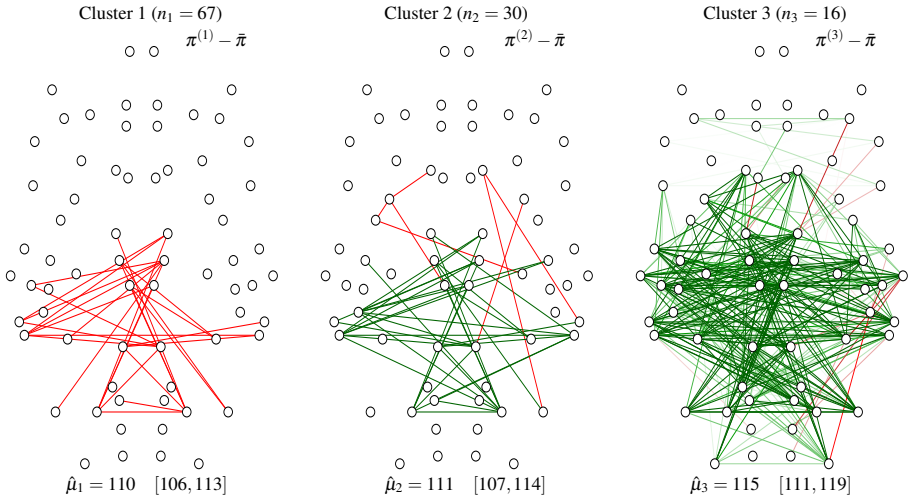

Fig. 2 For the three clusters, estimated deviation of the cluster-specific edge probability vector $\pi^{(h)}$ from the averaged edge probability vector $\bar{\pi} = \sum_{h=1}^{H} \nu_h \pi^{(h)}$, excluding insignificant changes (absolute difference less than 0.25) to improve visualization. Red correspond to negative deviations and green to positive ones. For each cluster is also indicated the total number of subjects and the posterior mean of the corresponding intercept μ_h in the response model along with its posterior quartiles. Nodes position corresponds to their spatial coordinates in the brain.

a reasonable scenario characterized by groups with low, averaged and high intelligence. We consider 5,000 MCMC chains, holding out the first 1,000 as burn-in.

Figure 2 shows estimated network deviations for each cluster, while displaying the three plots in increasing stochastic order of the posterior distribution for their corresponding offsets in the response model. Relating brain network deviations to their corresponding offsets in the response model provides interesting information on the relation between brain network structure and intelligence. In particular, we notice how positive location shifts in the posterior distribution for the offsets for the FSIQ are accompanied by increasingly likely connections between brain regions located in several anatomical lobes covering frontal, parietal, temporal, occipital and limbic. These deviations are particularly evident in the third group, which is characterized by the most evident positive shift in the conditional expectation for the FSIQ response. There are instead less differences in the posterior distribution for the offsets in the first and second group. It is, however, interesting to note how even a minor positive shift is already accompanied by an increased brain network connectivity mostly in parietal and temporal regions. These results are consistent with several previous neuroscience studies relating brain network to intelligence; see e.g. Colom et al. (2010) and the references cited therein.

References

Bigelow, J. L., and Dunson, D. B.: Bayesian semiparametric joint models for functional predictors. Journal of the American Statistical Association **104**, 26–36 (2009).

Colom, R., Karama, S., Jung, R. E. and Haier, R. J.: Human intelligence and brain networks. Dialogues in Clinical NeuroSciences **12**, 489–501 (2010).

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., and Killiany, R. J.: An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. NeuroImage **3**, 968–980 (2006).

Durante, D., Dunson, D. B. and Vogelstein, J. T.: Nonparametric Bayes Modeling of Populations of Networks. arXiv:1406.7851 (2015).

Durante, D. and Dunson, D. B.: Bayesian dynamic financial networks with time-varying predictors. Statistics & Probability Letters **93**, 19–26 (2014).

Erdös, P. and Rényi, A.: On random graphs. Publicationes Mathematicae Debrecen **5**, 290–297 (1959).

Frank, O. and Strauss, D.: Markov graphs. Journal of the American Statistical Association **81**, 832–842 (1986).

Hoff, P. D., Raftery, A. E. and Handcock, M. S.: Latent space approaches to social network analysis. Journal of the American Statistical Association **97**, 1090–1098 (2002).

Holland, P. W. and Leinhardt, S.: An exponential family of probability distributions for directed graphs. Journal of the American Statistical Association **76**, 33–50 (1981).

Koriakin, T. A., Mccurdy, M. D., Papazoglou, A., Pritchard, A. E., Zabel, T. A., Mahone, E. M., and Jacobson, L. A.: Classification of intellectual disability using the Wechsler Intelligence Scale for Children: Full Scale IQ or General Abilities Index? Developmental Medicine & Child Neurology **55**, 840–845 (2013).

Nowicki, N. and Snijders, T. A. B. Estimation and prediction for stochastic blockstructures. Journal of the American Statistical Association **96**, 1077–1087 (2001).

Roncal, W. G., Koterba, Z. H., Mhembere, D., Kleissas, D. M., Vogelstein, J. T., Burns, R., Bowles, A. R., Donavos, D. K., Ryman, S., Jung, R. E., Wu, L., Calhoun, V., and Vogelstein, R. J.: Migraine: MRI graph reliability analysis and inference for connectomics. IEEE Global Conference on Signal and Information Processing. IEEE (2013).

Snijders, T. A. B.: Models for longitudinal network data. In P. Carrington, J. Scott, and S. Wasserman, editors. Models and Methods in Social Network Analysis, pages 215–247. Cambridge University Press (2005).

Snijders, T. A. B., Pattison, E. P., Robins, G. L. and Handcock, M. S.: New specifications for exponential random graph models. Sociological Methodology **36**, 99–153 (2006)

Stephens, M.: Dealing with label switching in mixture models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **62**, 795–809 (2000).