Bayesian Analysis (2018) **13**, Number 1, pp. 29–58

Bayesian Inference and Testing of Group Differences in Brain Networks

Daniele Durante* and David B. Dunson†

Abstract. Network data are increasingly collected along with other variables of interest. Our motivation is drawn from neurophysiology studies measuring brain connectivity networks for a sample of individuals along with their membership to a low or high creative reasoning group. It is of paramount importance to develop statistical methods for testing of global and local changes in the structural interconnections among brain regions across groups. We develop a general Bayesian procedure for inference and testing of group differences in the network structure, which relies on a nonparametric representation for the conditional probability mass function associated with a network-valued random variable. By leveraging a mixture of low-rank factorizations, we allow simple global and local hypothesis testing adjusting for multiplicity. An efficient Gibbs sampler is defined for posterior computation. We provide theoretical results on the flexibility of the model and assess testing performance in simulations. The approach is applied to provide novel insights on the relationships between human brain networks and creativity.

Keywords: brain network, mixture model, multiple testing, nonparametric Bayes.

1 Introduction

There has been an increasing focus on using neuroimaging technologies to better understand the neural pathways underlying human behavior, abilities and neuropsychiatric diseases. The primary emphasis has been on relating the level of activity in brain regions to phenotypes. Activity measures are available via electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) – among others – and the aim is to produce a spatial map of the locations in the brain across which activity levels display evidence of change with the phenotype (e.g. Genovese et al., 2002; Tansey et al., 2014).

Although the above analyses remain an active area of research, more recently there has been a paradigm shift in neuroscience away from the modular approach, and towards studying brain connectivity networks and their relationship with phenotypes (Fuster, 2000, 2006). It has been increasingly realized that it is naive to study region-specific activity in isolation, and the overall circuit structure across the brain is a more important predictor of phenotypes (Bressler and Menon, 2010). Brain connectivity data are now available to facilitate this task, with non-invasive imaging technologies providing accurate brain network data at increasing spatial resolution; see Stirling and Elliott (2008), Craddock et al. (2013) and Wang et al. (2014) for an overview and recent developments.

*University of Padova, Department of Statistical Sciences. Via Cesare Battisti, 241, 35121 Padova, Italy, durante@stat.unipd.it

†Duke University, Department of Statistical Science. Box 90251, Durham, NC 27708-0251, USA, dunson@duke.edu

 https://doi.org/10.1214/16-BA1030

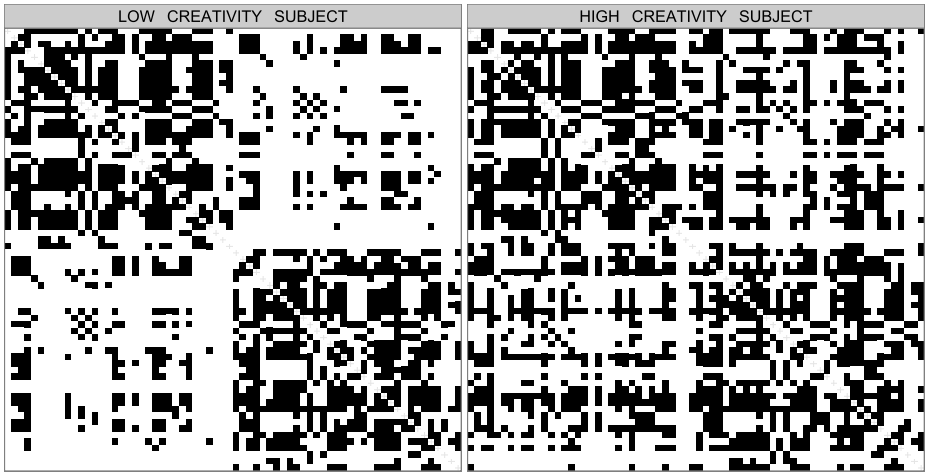

Figure 1: Adjacency matrices $\boldsymbol{A}_i$ representing the brain network of two subjects in the low and high creativity group. Black refers to an edge and white to a non-edge.

A common approach for constructing brain network data is based on the covariance in activity across brain regions estimated from fMRI data. For example, one can create a functional connectivity network from the inverse covariance matrix, with low values of the precision matrix suggesting evidence of conditional independence between pairs of brain regions (e.g. Ramsey et al., 2010; Smith et al., 2011; Simpson et al., 2013). Even if functional connectivity networks are of fundamental interest, the recent developments in diffusion tensor imaging (DTI) technologies (Craddock et al., 2013) have motivated an increasing focus on structural brain network data measuring anatomical connections made by axonal pathways.

DTI maps the diffusion of water molecules across biological tissues, thereby providing a better candidate to estimate axonal pathways. As the directional diffusion of the water within the brain tends to occur along the white matter fibers, current connectome preprocessing pipelines (e.g. Craddock et al., 2013; Gray Roncal et al., 2013) can produce an adjacency matrix $\boldsymbol{A}_i$ for each individual $i = 1, \ldots, n$, with elements $A_{i[vu]} = A_{i[uv]} = 1$ if there is at least one white matter fiber connecting brain regions $v = 2, \ldots, V$ and $u = 1, \ldots, v-1$ in individual i, and $A_{i[vu]} = A_{i[uv]} = 0$ otherwise. In our applications $V = 68$ and each node in the network characterizes a specific anatomical brain region according to the Desikan atlas (Desikan et al., 2006), with the first 34 in the left hemisphere and the remaining 34 in the right; see Figure 1 for an illustration. Refer also to Sporns (2013) for a discussion on functional and structural connectivity networks.

1.1 Motivating application and relevant literature

Recent studies provide brain networks along with a categorical variable. Examples include presence or absence of a neuropsychiatric disease, cognitive trait and rest-stimulus

states. There is a need for methods assessing how the brain connectivity structure varies across groups. We are specifically interested in studying the relationship between the brain connectivity structure and creative reasoning. For each individual $i = 1, \ldots, n$, data consist of an indicator of creative reasoning y_i and an adjacency matrix $\boldsymbol{A}_i$ representing the undirected structural brain network. We focus on dataset MRN-111 available at `http://openconnecto.me/data/public/MR/`, preselecting subjects having low (< 90, $y_i = 1$) or high (> 111, $y_i = 2$) creative reasoning scores. The first group comprises 17 subjects and the second 19, with thresholds chosen to correspond to the 0.15 and 0.85 quantiles. Creativity scores are measured via the composite creativity index (CCI) (Jung et al., 2010). We are interested in assessing evidence of differences in brain connectivity between the low and high creativity groups, while additionally inferring the types of differences and learning which connections are responsible for these variations. Note that we are not attempting to estimate a network – as in graphical modeling – but we are focused on testing of differences between groups in network-valued data.

Flexible statistical methods for analyzing brain networks have lagged behind the increasingly routine collection of such data in neuroscience. A major barrier to progress in this area is that the development of statistical methodologies for formal and robust inference on network data is a challenging task. Networks represent a type of object data – a concept encompassing a broad class of non-standard data types, ranging from functions to images and trees; refer to Wang and Marron (2007) and the references cited therein for an overview. Such data require adaptations of classical modeling frameworks to non-standard spaces. This is particularly true for inference on network data in which the set of methodologies and concepts required to test for changes in underlying connectivity structures is necessarily distinct from standard data analysis strategies.

There has been some emphasis in the literature on developing methods for addressing our goals; see Bullmore and Sporns (2009), Stam (2014) and the references cited therein. The main focus is on reducing each observed network $\boldsymbol{A}_i$, $i = 1, \ldots, n$ to a vector of summary statistics $\boldsymbol{\theta}_i = (\theta_{i1}, \ldots, \theta_{ip})^{\mathrm{T}}$ and then apply standard procedures, such as the multivariate analysis of variance (MANOVA), to test for changes in these vectors across groups. Summary statistics are commonly chosen to represent global network characteristics of interest, such as the number of connections, the average path length and the clustering coefficient (Rubinov and Sporns, 2010). Similar procedures have been recently employed in exploring the relationship between the brain network and neuropsychiatric diseases, such as Parkinson's (Olde Dubbelink et al., 2014) and Alzheimer's (Daianu et al., 2013), but the analyses are sensitive to the chosen network topological measures, with substantially different results obtained for different types of summary statistics. Simpson et al. (2011) and Simpson et al. (2012) improve choice of network summary statistics via a data driven procedure which exploits exponential random graph models (e.g. Frank and Strauss, 1986; Wasserman and Pattison, 1996) and related validation procedures (Hunter et al., 2008a,b) to detect the topological measures that better characterize the observed networks. Although this is a valuable procedure, inference is still available only on the scale of the network summary statistics, which typically discard important information about the brain connectivity architecture that may crucially explain differences across groups. Refer to Arden et al. (2010) for a review on inconsistencies in results relating brain connectivity networks to creative reasoning.

An alternative approach is to avoid discarding information by separately testing for differences between groups in each edge probability, while adjusting the significance threshold for multiple testing via false discovery rate (FDR) control. As there are $V(V-1)/2$ pairs of brain regions under study – with $V = 68$ using the Desikan atlas (Desikan et al., 2006) – the number of tests is substantial. Such massively univariate approaches do not incorporate network information, leading to low power (Fornito et al., 2013), and underestimating the variations of the brain connections across groups. Recent proposals try to gain power by replacing the common Benjamini and Hochberg (1995) approach with thresholding procedures that account for the network structure in the data (Zalesky et al., 2010). However, such approaches require careful interpretation, while being highly computationally intensive, requiring permutation testing and choice of suprathreshold links. Instead of controlling FDR thresholds, Scott et al. (2015) gain power in multiple testing by using auxiliary data – such as spatial proximity – to inform the posterior probability that specific pairs of nodes interact differently across groups or with respect to a baseline. Ginestet et al. (2014) focus instead on assessing evidence of global changes in the brain structure by testing for group differences in the expected Laplacians.

Scott et al. (2015) and Ginestet et al. (2014) substantially improve state of the art in local and global hypothesis testing for network data, respectively, but are characterized by a similar key issue, motivating our methodology. Specifically, previous procedures test for changes across groups in marginal (Scott et al., 2015) or expected (Ginestet et al., 2014) structures associated with the network-valued random variable, and hence cannot detect variations in the probabilistic generative mechanism that go beyond their focus. Similarly to much simpler settings, substantially different joint probability mass functions (pmf) for a network-valued random variable can have equal expectation or induce the same marginal distributions – characterized by the edge probabilities. Hence, these procedures are expected to fail in scenarios where the changes in the network-valued random variable are due to variations in more complex functionals. Model misspecification can have a major effect on the quality of inference (Deegan, 1976; Begg and Lagakos, 1990; DiRienzo and Lagakos, 2001), providing biased and inaccurate conclusions.

1.2 Outline of our methodology

In order to avoid the issues discussed above, it is fundamental to define a statistical model which is sufficiently flexible to accurately approximate any probabilistic generative mechanism underlying the observed data. Durante et al. (2016) recently proposed a flexible mixture of low-rank factorizations to characterize the distribution of a network-valued random variable. We generalize their statistical model to allow the probabilistic generative mechanism associated with the brain networks to change across groups, without reducing data to summary measures prior to statistical analysis.

We accomplish the above goal by factorizing the joint pmf for the random variable generating data $(y_i, \boldsymbol{A}_i)$, $i = 1, \ldots, n$, as the product of the marginal pmf for the categorical predictor and the conditional pmf for the network-valued random variable given the group membership defined by the categorical predictor. By modeling the collection of group-dependent pmfs for the network-valued random variable via a flexible mixture of low-rank factorizations with group-specific mixing probabilities, we develop a simple

global test for assessing evidence of group differences in the entire distribution of the network-valued random variable, rather than focusing inference only on changes in selected functionals. Differently from Ginestet et al. (2014), our procedure additionally incorporates local testing for changes in edge probabilities across groups, in line with Scott et al. (2015) – which in turn do not consider global tests. By explicitly borrowing strength within the network via matrix factorizations, we substantially improve power in our multiple local tests compared to standard FDR control procedures.

In Section 2 we describe the model formulation, with a key focus on the associated testing procedures. Prior specification, theoretical properties and posterior computation are considered in Section 3. Section 4 provides simulations to assess inference and testing performance of our procedures. Results for our motivating neuroscience application are discussed in Section 5. Concluding remarks are provided in Section 6.

2 Model formulation and testing

2.1 Notation and motivation

Let $(y_i, \boldsymbol{A}_i)$ represent the creativity group and the undirected network observation, respectively, for subject $i = 1, \ldots, n$, with $y_i \in \mathbb{Y} = \{1, 2\}$ and $\boldsymbol{A}_i$ the $V \times V$ adjacency matrix characterizing the edges in the network. As the brain network structure is available via undirected edges and self-relationships are not of interest, we model $(y_i, \boldsymbol{A}_i)$ by focusing on the random variable $\{\mathcal{Y}, \mathcal{L}(\boldsymbol{\mathcal{A}})\}$ generating data $\{y_i, \mathcal{L}(\boldsymbol{A}_i)\}$ with $\mathcal{L}(\boldsymbol{A}_i) = (A_{i[21]}, A_{i[31]}, \ldots, A_{i[V1]}, A_{i[32]}, \ldots, A_{i[V2]}, \ldots, A_{i[V(V-1)]})^{\mathrm{T}} \in \mathbb{A}_V = \{0, 1\}^{V(V-1)/2}$ the vector encoding the lower triangular elements of $\boldsymbol{A}_i$, which uniquely define the network as $A_{i[vu]} = A_{i[uv]}$ for every $v = 2, \ldots, V$, $u = 1, \ldots, v-1$ and $i = 1, \ldots, n$.

Let $p_{\mathcal{Y},\mathcal{L}(\boldsymbol{\mathcal{A}})} = \{p_{\mathcal{Y},\mathcal{L}(\boldsymbol{\mathcal{A}})}(y, \boldsymbol{a}) : y \in \mathbb{Y}, \boldsymbol{a} \in \mathbb{A}_V\}$ denote the joint pmf for the random variable $\{\mathcal{Y}, \mathcal{L}(\boldsymbol{\mathcal{A}})\}$ with $p_{\mathcal{Y},\mathcal{L}(\boldsymbol{\mathcal{A}})}(y, \boldsymbol{a}) = \mathrm{pr}\{\mathcal{Y} = y, \mathcal{L}(\boldsymbol{\mathcal{A}}) = \boldsymbol{a}\}$, $y \in \mathbb{Y}$ and $\boldsymbol{a} \in \mathbb{A}_V$ a network configuration. Assessing evidence of global association between $\mathcal{Y}$ and $\mathcal{L}(\boldsymbol{\mathcal{A}})$ – under the above notation – formally requires testing the global null hypothesis

$$H_0 : p_{\mathcal{Y},\mathcal{L}(\boldsymbol{\mathcal{A}})}(y, \boldsymbol{a}) = p_{\mathcal{Y}}(y) p_{\mathcal{L}(\boldsymbol{\mathcal{A}})}(\boldsymbol{a}), \tag{1}$$

for all $y \in \mathbb{Y}$ and $\boldsymbol{a} \in \mathbb{A}_V$, versus the alternative

$$H_1 : p_{\mathcal{Y},\mathcal{L}(\boldsymbol{\mathcal{A}})}(y, \boldsymbol{a}) \neq p_{\mathcal{Y}}(y) p_{\mathcal{L}(\boldsymbol{\mathcal{A}})}(\boldsymbol{a}), \tag{2}$$

for some $y \in \mathbb{Y}$ and $\boldsymbol{a} \in \mathbb{A}_V$, where $p_{\mathcal{Y}}(y) = \mathrm{pr}(\mathcal{Y} = y)$, $y \in \mathbb{Y}$ characterizes the marginal pmf of the grouping variable, whereas $p_{\mathcal{L}(\boldsymbol{\mathcal{A}})}(\boldsymbol{a}) = \mathrm{pr}\{\mathcal{L}(\boldsymbol{\mathcal{A}}) = \boldsymbol{a}\}$, $\boldsymbol{a} \in \mathbb{A}_V$ denotes the unconditional pmf for the network-valued random variable. The system of hypotheses (1)–(2) assesses evidence of global changes in the entire probability mass function, rather than on selected functionals or summary statistics, and hence is more general than Ginestet et al. (2014) and joint tests on network measures.

Recalling our neuroscience application, rejection of H_0 implies that there are differences in the brain architecture across creativity groups, but fails to provide insights on the reasons for these variations. Global differences may be attributable to several

underlying mechanisms, including changes in specific interconnection circuits. As discussed in Section 1, local testing of group differences in the edge probabilities is of key interest in neuroscience applications in highlighting which brain connection measurements $\mathcal{L}(\mathcal{A})_l \in \{0,1\}$, $l = 1, \ldots, V(V-1)/2$ – characterizing the marginals of $\mathcal{L}(\boldsymbol{\mathcal{A}})$ – are potentially responsible for the global association between $\mathcal{Y}$ and $\mathcal{L}(\boldsymbol{\mathcal{A}})$. Hence, consistently with these interests, we also incorporate in our analyses the multiple local tests assessing – for each pair $l = 1, \ldots, V(V-1)/2$ – evidence against the null hypothesis of independence between $\mathcal{L}(\mathcal{A})_l$ and $\mathcal{Y}$

$$H_{0l} : p_{\mathcal{Y},\mathcal{L}(\mathcal{A})_l}(y, a_l) = p_{\mathcal{Y}}(y) p_{\mathcal{L}(\mathcal{A})_l}(a_l), \tag{3}$$

for all $y \in \mathbb{Y}$ and $a_l \in \{0,1\}$, versus the alternative

$$H_{1l} : p_{\mathcal{Y},\mathcal{L}(\mathcal{A})_l}(y, a_l) \neq p_{\mathcal{Y}}(y) p_{\mathcal{L}(\mathcal{A})_l}(a_l), \tag{4}$$

for some $y \in \mathbb{Y}$ and $a_l \in \{0,1\}$. In hypotheses (3)–(4), the quantity $p_{\mathcal{Y},\mathcal{L}(\mathcal{A})_l}(y, a_l)$ denotes $\text{pr}\{\mathcal{Y} = y, \mathcal{L}(\mathcal{A})_l = a_l\}$, while $p_{\mathcal{L}(\mathcal{A})_l}(a_l) = \text{pr}\{\mathcal{L}(\mathcal{A})_l = a_l\}$.

In order to develop robust methods to test the global system (1)–(2), and the multiple locals (3)–(4), it is fundamental to define a representation for $p_{\mathcal{Y},\mathcal{L}(\boldsymbol{\mathcal{A}})}$ which is provably flexible in approximating any joint pmf for the data $\{y_i, \mathcal{L}(\boldsymbol{A}_i)\}$, $i = 1, \ldots, n$. As $\mathcal{L}(\boldsymbol{\mathcal{A}})$ is a highly multidimensional variable on a non-standard space, we additionally seek to reduce dimensionality in characterizing $p_{\mathcal{Y},\mathcal{L}(\boldsymbol{\mathcal{A}})}$, while looking for a representation which facilitates simple derivation of $p_{\mathcal{Y},\mathcal{L}(\mathcal{A})_l}(y, a_l)$ and $p_{\mathcal{L}(\mathcal{A})_l}(a_l)$ from $p_{\mathcal{Y},\mathcal{L}(\boldsymbol{\mathcal{A}})}$.

2.2 Dependent mixture of low-rank factorizations

According to the goals described above, we start by factorizing $p_{\mathcal{Y},\mathcal{L}(\boldsymbol{\mathcal{A}})}$ as

$$p_{\mathcal{Y},\mathcal{L}(\boldsymbol{\mathcal{A}})}(y, \boldsymbol{a}) = p_{\mathcal{Y}}(y) p_{\mathcal{L}(\boldsymbol{\mathcal{A}})|y}(\boldsymbol{a}) = \text{pr}(\mathcal{Y} = y)\text{pr}\{\mathcal{L}(\boldsymbol{\mathcal{A}}) = \boldsymbol{a} \mid \mathcal{Y} = y\}, \tag{5}$$

for every $y \in \mathbb{Y}$ and $\boldsymbol{a} \in \mathbb{A}_V$. It is always possible to define the joint probability mass function $p_{\mathcal{Y},\mathcal{L}(\boldsymbol{\mathcal{A}})}$ as the product of the marginal pmf $p_{\mathcal{Y}} = \{p_{\mathcal{Y}}(y) : y \in \mathbb{Y}\}$ for the grouping variable and the conditional pmfs $p_{\mathcal{L}(\boldsymbol{\mathcal{A}})|y} = \{p_{\mathcal{L}(\boldsymbol{\mathcal{A}})|y}(\boldsymbol{a}) : \boldsymbol{a} \in \mathbb{A}_V\}$ for the network-valued random variable given the group $y \in \mathbb{Y}$. This also favors inference on how the network structure varies across the two groups, with $p_{\mathcal{L}(\boldsymbol{\mathcal{A}})|1}$ and $p_{\mathcal{L}(\boldsymbol{\mathcal{A}})|2}$ fully characterizing such variations. Although we treat $\mathcal{Y}$ as a random variable through a prospective likelihood, our methodology remains also valid for studies that sample the groups under a retrospective design.

Under factorization (5), the global test (1)–(2) coincides with assessing whether the conditional pmf of the network-valued random variable remains equal or shifts across the two groups. Hence, under (5), the hypotheses (1)–(2) reduce to

$$H_0 : p_{\mathcal{L}(\boldsymbol{\mathcal{A}})|1}(\boldsymbol{a}) = p_{\mathcal{L}(\boldsymbol{\mathcal{A}})|2}(\boldsymbol{a}), \quad \text{for all } \boldsymbol{a} \in \mathbb{A}_V, \tag{6}$$

versus the alternative

$$H_1 : p_{\mathcal{L}(\boldsymbol{\mathcal{A}})|1}(\boldsymbol{a}) \neq p_{\mathcal{L}(\boldsymbol{\mathcal{A}})|2}(\boldsymbol{a}), \quad \text{for some } \boldsymbol{a} \in \mathbb{A}_V. \tag{7}$$

In order to develop a provably general and robust strategy to test (6)–(7) the key challenge relies in flexibly modeling the conditional pmfs $p_{\mathcal{L}(\boldsymbol{\mathcal{A}})|1}$ and $p_{\mathcal{L}(\boldsymbol{\mathcal{A}})|2}$ characterizing the distribution of the network-valued random variable in the first and second group, respectively. In fact, for every group $y \in \mathbb{Y}$, one needs a parameter $p_{\mathcal{L}(\boldsymbol{\mathcal{A}})|y}(\boldsymbol{a})$ for every possible network configuration $\boldsymbol{a} \in \mathbb{A}_V$ to uniquely characterize $p_{\mathcal{L}(\boldsymbol{\mathcal{A}})|y}$, with the number of configurations being $|\mathbb{A}_V| = 2^{V(V-1)/2}$. For example, in our neuroscience application $|\mathbb{A}_{68}| = 2^{68(68-1)/2} - 1 = 2^{2,278} - 1$ free parameters are required to uniquely define the pmf of the brain networks in each group $y \in \mathbb{Y}$, under the usual restriction $\sum_{\boldsymbol{a} \in \mathbb{A}_{68}} p_{\mathcal{L}(\boldsymbol{\mathcal{A}})|y}(\boldsymbol{a}) = 1$. Clearly this number of parameters to test is massively larger than the sample size available in neuroscience applications. Hence, to facilitate tractable testing procedures it is necessary to substantially reduce dimensionality. However, in reducing dimension, it is important to avoid making overly restrictive assumptions that lead to formulations sensitive to issues arising from model misspecification.

Focused on modeling the pmf $p_{\mathcal{L}(\boldsymbol{\mathcal{A}})}$ of a network-valued random variable, without considering hypothesis tests or additional data on a categorical predictor, Durante et al. (2016) proposed a mixture of low-rank factorizations which reduces dimensionality by exploiting network information, but maintains flexibility. Although this provides an appealing building block for our testing procedures, global and local testing, and inference on group differences, are not straightforward adds on to their approach. As a first step towards constructing our tests, we generalize their model to allow group differences via

$$p_{\mathcal{L}(\boldsymbol{\mathcal{A}})|y}(\boldsymbol{a}) = \text{pr}\{\mathcal{L}(\boldsymbol{\mathcal{A}}) = \boldsymbol{a} \mid \mathcal{Y} = y\} = \sum_{h=1}^{H} \nu_{hy} \prod_{l=1}^{V(V-1)/2} (\pi_l^{(h)})^{a_l} (1 - \pi_l^{(h)})^{1-a_l}, \qquad (8)$$

for each configuration $\boldsymbol{a} \in \mathbb{A}_V$ and group $y \in \{1, 2\}$, with the edge probability vectors $\boldsymbol{\pi}^{(h)} = (\pi_1^{(h)}, \ldots, \pi_{V(V-1)/2}^{(h)})^{\text{T}} \in (0,1)^{V(V-1)/2}$ in each mixture component h, defined as

$$\boldsymbol{\pi}^{(h)} = \left\{1 + \exp(-\boldsymbol{Z} - \boldsymbol{D}^{(h)})\right\}^{-1}, \quad \boldsymbol{D}^{(h)} = \mathcal{L}(\boldsymbol{X}^{(h)} \boldsymbol{\Lambda}^{(h)} \boldsymbol{X}^{(h)\text{T}}), \quad h = 1, \ldots, H, \qquad (9)$$

with $\boldsymbol{X}^{(h)} \in \Re^{V \times R}$, $\boldsymbol{\Lambda}^{(h)}$ diagonal with R non-negative weights $\lambda_1^{(h)}, \ldots, \lambda_R^{(h)}$, and $\boldsymbol{Z} \in \Re^{V(V-1)/2}$. Representation (8) defines $p_{\mathcal{L}(\boldsymbol{\mathcal{A}})|y}$ via a flexible dependent mixture model which borrows strength across groups in characterizing the shared mixture components, while allowing flexible modeling of the conditional pmfs $p_{\mathcal{L}(\boldsymbol{\mathcal{A}})|y}$ via group-specific mixing probabilities $\boldsymbol{\nu}_y = (\nu_{1y}, \ldots, \nu_{Hy})$, $y \in \{1, 2\}$, with $\nu_{hy} \in (0, 1)$ for all $h = 1, \ldots, H$ and $\sum_{h=1}^{H} \nu_{hy} = 1$ for every $y \in \{1, 2\}$. In (9) the logistic mapping is applied element-wise.

In order to reduce dimensionality and efficiently borrow information within the network, the characterization of the mixture components in (9) adapts concepts from the literature on latent variable modeling of networks. Refer to Nowicki and Snijders (2001), Airoldi et al. (2008), Hoff et al. (2002) and Hoff (2008) for popular specifications in modeling of a single network observation. Within each mixture component, connections among pairs of nodes are characterized as conditionally independent Bernoulli random variables given their component-specific edge probabilities $\pi_l^{(h)}$, $l = 1, \ldots, V(V-1)/2$, with these probabilities further characterized as a function of node-specific latent variables. In particular, we define each component-specific log-odds vector as the sum of a

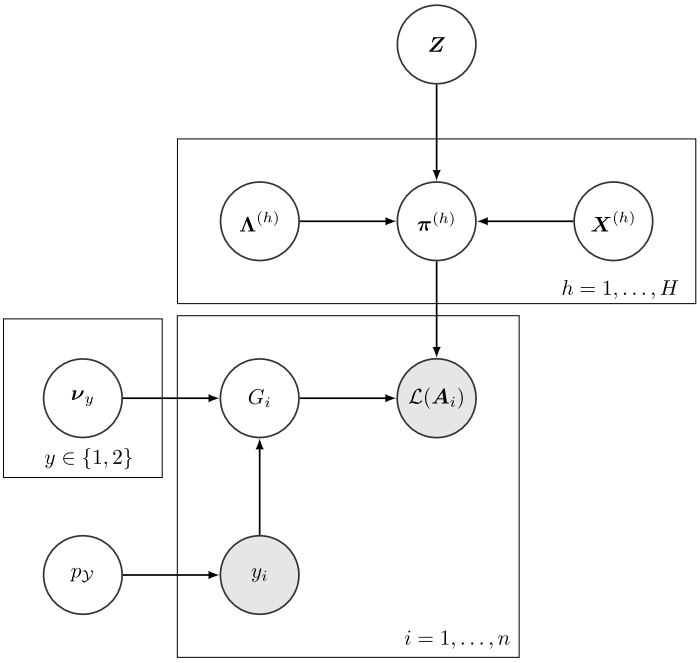

Figure 2: Graphical representation of the mechanism to generate data $\{y_i, \mathcal{L}(\boldsymbol{A}_i)\}$, $i = 1, \ldots, n$, under representation (5) and (8)–(9) for the joint pmf $p_{\mathcal{Y},\mathcal{L}(\boldsymbol{\mathcal{A}})}$.

shared similarity $\boldsymbol{Z} \in \Re^{V(V-1)/2}$, and a component-specific one $\boldsymbol{D}^{(h)} \in \Re^{V(V-1)/2}$ arising from the weighted dot product of node-specific latent coordinate vectors defining the rows of the $V \times R$ – typically $R \ll V$ – matrix $\boldsymbol{X}^{(h)}$, for $h = 1, \ldots, H$. In fact, letting l denote the pair of nodes v and u, $v > u$, under (9), the probability of an edge between v and u in component h increases with Z_l and $\mathcal{L}(X^{(h)}\Lambda^{(h)}X^{(h)\mathrm{T}})_l = \sum_{r=1}^{R} \lambda_r^{(h)} X_{vr}^{(h)} X_{ur}^{(h)}$. Representation (9) provides an over-complete factorization – a common approach providing several benefits in Bayesian hierarchical modeling of multidimensional data (e.g. Bhattacharya and Dunson, 2011; Ghosh and Dunson, 2009). In fact, factorization (9) is appealing in reducing dimensionality, accommodating topological network properties (Hoff, 2008) and improving mixing performance (Gelman et al., 2008). Our focus is on using the resulting flexible and tractable formulation (8)–(9) to draw inference on changes in identified functionals of interest arising from the pmf of our network-valued random variable and develop robust procedures for global and local testing.

Figure 2 outlines the mechanism to generate the data $\{y_i, \mathcal{L}(\boldsymbol{A}_i)\}$ from the random variable $\{\mathcal{Y}, \mathcal{L}(\boldsymbol{\mathcal{A}})\}$ with pmf factorized as in (5) and (8)–(9). According to Figure 2, the indicator group y_i is sampled from $p_{\mathcal{Y}}$. The network $\mathcal{L}(\boldsymbol{A}_i)$ is instead generated conditioned on y_i under the mixture representation in (8). In particular, given $y_i = y$, we first choose a mixture component by sampling the latent indicator $G_i \in \{1, \ldots, H\}$ with conditional pmf defined by the mixing probabilities, so that $p_{G_i|y}(h) = \nu_{hy}$. Then,

given $G_i = h$ and the corresponding edge probability vector $\boldsymbol{\pi}^{(h)}$ – factorized as in (9) – the network $\mathcal{L}(\boldsymbol{A}_i)$ is generated by sampling its edges $\mathcal{L}(A_i)_l$, $l = 1, \ldots, V(V-1)/2$ from conditionally independent Bernoulli variables. Hence, the dependence on the groups is introduced in the assignments to the mixture components via group-specific mixing probabilities, so that brain networks in the same component share a common edge probability vector $\boldsymbol{\pi}^{(h)}$, with the probability assigned to each component changing across the two groups. This simple generative mechanism is appealing in facilitating tractable posterior computation and inference.

A key aspect in representation (8)–(9) is that it allows dimensionality reduction, while preserving flexibility. As stated in Proposition 1, such a representation is sufficiently flexible to define any collection of group-dependent pmfs $p_{\mathcal{L}(\boldsymbol{\mathcal{A}})|1}$, $p_{\mathcal{L}(\boldsymbol{\mathcal{A}})|2}$.

Proposition 1. *Any collection of group-dependent probability mass functions* $p_{\mathcal{L}(\boldsymbol{\mathcal{A}})|y} \in \mathcal{P}_{|\mathbb{A}_V|} = \{p_{\mathcal{L}(\boldsymbol{\mathcal{A}})|y} : 0 \leq p_{\mathcal{L}(\boldsymbol{\mathcal{A}})|y}(\boldsymbol{a}) \leq 1$ *for all* $\boldsymbol{a} \in \mathbb{A}_V, \sum_{\boldsymbol{a}\in\mathbb{A}_V} p_{\mathcal{L}(\boldsymbol{\mathcal{A}})|y}(\boldsymbol{a}) = 1\}$, $y \in \{1, 2\}$ *can be characterized as in* (8) *for some* H, *with component-specific edge probability vectors* $\boldsymbol{\pi}^{(h)}$, $h = 1, \ldots, H$ *factorized as in* (9) *for some* R.

This additionally ensures that any joint probability mass function $p_{\mathcal{Y},\mathcal{L}(\boldsymbol{\mathcal{A}})}$ for the random variable $\{\mathcal{Y}, \mathcal{L}(\boldsymbol{\mathcal{A}})\}$ admits representation (5), (8)–(9) and hence our formulation can be viewed as fully general and robust against model misspecification in testing (6)–(7), given sufficiently flexible priors for the components. See the online supplementary materials (Durante and Dunson, 2016) for proofs.

2.3 Global and local testing under the proposed statistical model

Including group dependence only in the mixing probabilities favors borrowing of information across the groups in modeling $\boldsymbol{\pi}^{(h)}$, $h = 1, \ldots, H$, while massively reducing the number of parameters to test in (6)–(7) from $2\{2^{V(V-1)/2} - 1\}$ to $2(H-1)$. In fact, the characterization of $p_{\mathcal{L}(\boldsymbol{\mathcal{A}})|y}$ in (8)–(9) further simplifies the system (6)–(7) to only testing the equality of the group-specific mixing probability vectors

$$H_0 : (\nu_{11}, \ldots, \nu_{H1}) = (\nu_{12}, \ldots, \nu_{H2}) \quad \text{versus} \quad H_1 : (\nu_{11}, \ldots, \nu_{H1}) \neq (\nu_{12}, \ldots, \nu_{H2}). \tag{10}$$

Recalling Proposition 1, under our formulation, the system (10) uniquely characterizes the global hypotheses (1)–(2).

In developing methodologies for the multiple local tests in (3)–(4) under our model formulation, we measure the association between $\mathcal{L}(\mathcal{A})_l$ and $\mathcal{Y}$ by exploiting the model-based version of the Cramer's V proposed in Dunson and Xing (2009), obtaining

$$\begin{aligned} \rho_l^2 &= \frac{1}{\min\{2,2\}-1} \sum_{y=1}^{2} \sum_{a_l=0}^{1} \frac{\{p_{\mathcal{Y},\mathcal{L}(\mathcal{A})_l}(y, a_l) - p_{\mathcal{Y}}(y) p_{\mathcal{L}(\mathcal{A})_l}(a_l)\}^2}{p_{\mathcal{Y}}(y) p_{\mathcal{L}(\mathcal{A})_l}(a_l)} \\ &= \sum_{y=1}^{2} \sum_{a_l=0}^{1} \frac{\{p_{\mathcal{Y}}(y) p_{\mathcal{L}(\mathcal{A})_l|y}(a_l) - p_{\mathcal{Y}}(y) p_{\mathcal{L}(\mathcal{A})_l}(a_l)\}^2}{p_{\mathcal{Y}}(y) p_{\mathcal{L}(\mathcal{A})_l}(a_l)} \end{aligned}$$

$$= \sum_{y=1}^{2} p_{\mathcal{Y}}(y) \sum_{a_l=0}^{1} \frac{\left\{p_{\mathcal{L}(\mathcal{A})_l|y}(a_l) - p_{\mathcal{L}(\mathcal{A})_l}(a_l)\right\}^2}{p_{\mathcal{L}(\mathcal{A})_l}(a_l)}. \tag{11}$$

Measuring the local association with $\rho_l \in (0,1)$ provides an appealing choice in terms of interpretation, with $\rho_l = 0$ meaning that $p_{\mathcal{Y},\mathcal{L}(\mathcal{A})_l}(y, a_l) = p_{\mathcal{Y}}(y)p_{\mathcal{L}(\mathcal{A})_l}(a_l)$, for all $y \in \mathbb{Y}$ and $a_l \in \{0,1\}$, and hence the random variable $\mathcal{L}(\mathcal{A})_l$ modeling the presence or absence of an edge among the lth pair of nodes, has no differences across groups. Beside incorporating a fully general and tractable global test, our model formulation is particularly appealing also in addressing issues associated with local multiple testing in the network framework. First, as stated in Proposition 2, each ρ_l, $l = 1, \ldots, V(V-1)/2$, can be easily computed from the quantities in our model.

Proposition 2. *Based on factorizations* (5) *and* (8), $p_{\mathcal{L}(\mathcal{A})_l|y}(1) = 1 - p_{\mathcal{L}(\mathcal{A})_l|y}(0) = \sum_{h=1}^{H} \nu_{hy}\pi_l^{(h)}$, *and* $p_{\mathcal{L}(\mathcal{A})_l}(1) = 1 - p_{\mathcal{L}(\mathcal{A})_l}(0) = \sum_{y=1}^{2} p_{\mathcal{Y}}(y) \sum_{h=1}^{H} \nu_{hy}\pi_l^{(h)}$.

Second, the shared dependence on a common set of node-specific latent coordinates characterizing the construction of the edge probability vector $\boldsymbol{\pi}^{(h)}$ within each mixture component $h = 1, \ldots, H$ in (9), explicitly accounts for specific dependence structures in brain connections. According to Hoff (2008), factorization (9) can accurately accommodate key topological properties including block structures, homophily behaviors and transitive edge patterns – among others. As a result – in line with Scott et al. (2015) – informing our local testing procedures about these structures, is expected to substantially improve power compared to standard FDR control procedures.

3 Prior specification and posterior computation

3.1 Prior specification and properties

We specify independent priors $p_{\mathcal{Y}} \sim \Pi_y$, $\boldsymbol{Z} = (Z_1, \ldots, Z_{V(V-1)/2})^{\mathrm{T}} \sim \Pi_Z$, $\boldsymbol{X}^{(h)} \sim \Pi_X$, $\boldsymbol{\lambda}^{(h)} = (\lambda_1^{(h)}, \ldots, \lambda_R^{(h)})^{\mathrm{T}} \sim \Pi_\lambda$, $h = 1, \ldots, H$ and $\boldsymbol{\nu}_y = (\nu_{1y}, \ldots, \nu_{Hy}) \sim \Pi_\nu$, $y \in \{1,2\}$, to induce a prior Π on the joint pmf $p_{\mathcal{Y},\mathcal{L}(\boldsymbol{\mathcal{A}})}$ with full support in $\mathcal{P}_{2\times|\mathbb{A}_V|} = \{p_{\mathcal{Y},\mathcal{L}(\boldsymbol{\mathcal{A}})} : 0 \leq p_{\mathcal{Y},\mathcal{L}(\boldsymbol{\mathcal{A}})}(y, \boldsymbol{a}) \leq 1$ for all $y \in \{1,2\}, \boldsymbol{a} \in \mathbb{A}_V$, with $\sum_{y\in\{1,2\},\boldsymbol{a}\in\mathbb{A}_V} p_{\mathcal{Y},\mathcal{L}(\boldsymbol{\mathcal{A}})}(y, \boldsymbol{a}) = 1\}$, while obtaining desirable asymptotic behavior, simple posterior computation and allowance for testing. Prior support is a key property to retain the flexibility associated with our statistical model and testing procedures, when performing posterior inference.

As $p_{\mathcal{Y}}$ is the pmf for a categorical variable on two levels, we let $1 - p_{\mathcal{Y}}(2) = p_{\mathcal{Y}}(1) \sim \text{Beta}(a, b)$, and consider the same prior specification suggested by Durante et al. (2016) for the quantities in (9) by choosing Gaussian priors for the entries in $\boldsymbol{Z}$, standard Gaussians for the elements in the coordinates matrix $\boldsymbol{X}^{(h)}$, and multiplicative inverse gammas for $\boldsymbol{\lambda}^{(h)} \sim \text{MIG}(a_1, a_2)$, $h = 1, \ldots, H$, (Bhattacharya and Dunson, 2011). This choice for Π_λ favors shrinkage, with elements in $\boldsymbol{\lambda}^{(h)}$ increasingly concentrated close to 0 as r increases, so as to shrink towards lower dimensional representations and adaptively penalize high dimensional ones. A key property of our prior specification is incorporation of global testing (10) in the definition of Π_ν. Specifically letting $\boldsymbol{\upsilon} = (\upsilon_1, \ldots, \upsilon_H)$ and

$\boldsymbol{\upsilon}_y = (\upsilon_{1y}, \ldots, \upsilon_{Hy})$, we induce Π_ν through

$$\begin{aligned} \boldsymbol{\nu}_y &= (1-T)\boldsymbol{\upsilon} + T\boldsymbol{\upsilon}_y, \quad y \in \{1,2\}, \\ \boldsymbol{\upsilon} &\sim \text{Dir}(1/H, \ldots, 1/H), \quad \boldsymbol{\upsilon}_y \sim \text{Dir}(1/H, \ldots, 1/H),\ y \in \{1,2\}, \\ T &\sim \text{Bern}\{\text{pr}(H_1)\}. \end{aligned} \tag{12}$$

In (12), T is a hypothesis indicator, with $T = 0$ for H_0 and $T = 1$ for H_1. Under H_1, we generate group-specific mixing probabilities independently, while under H_0 we have equal probability vectors. By choosing small values for the parameters in the Dirichlet priors, we favor automatic deletion of redundant components (Rousseau and Mengersen, 2011). In assessing evidence in favor of the alternative, we can rely on the posterior probability, $\text{pr}[H_1 \mid \{\boldsymbol{y}, \mathcal{L}(\boldsymbol{A})\}] = 1 - \text{pr}[H_0 \mid \{\boldsymbol{y}, \mathcal{L}(\boldsymbol{A})\}]$ which can be easily obtained from the output of the Gibbs sampler proposed below. Specifically, under prior (12), and exploiting the hierarchical structure of our dependent mixture model – summarized in Figure 2 – the full conditional $\text{pr}(T = 1 \mid -) = \text{pr}(H_1 \mid -) = 1 - \text{pr}(H_0 \mid -)$ is

$$\begin{aligned} \text{pr}(H_1 \mid -) &= \frac{\text{pr}(H_1) \prod_{y=1}^2 \int (\prod_{h=1}^H \upsilon_{hy}^{n_{hy}}) d\Pi_{\upsilon_y}}{\text{pr}(H_0) \int (\prod_{h=1}^H \upsilon_h^{n_h}) d\Pi_\upsilon + \text{pr}(H_1) \prod_{y=1}^2 \int (\prod_{h=1}^H \upsilon_{hy}^{n_{hy}}) d\Pi_{\upsilon_y}} \\ &= \frac{\text{pr}(H_1) \prod_{y=1}^2 \{\text{B}(\boldsymbol{\alpha} + \bar{\boldsymbol{n}}_y)/\text{B}(\boldsymbol{\alpha})\}}{\text{pr}(H_0)\text{B}(\boldsymbol{\alpha} + \bar{\boldsymbol{n}})/\text{B}(\boldsymbol{\alpha}) + \text{pr}(H_1) \prod_{y=1}^2 \{\text{B}(\boldsymbol{\alpha} + \bar{\boldsymbol{n}}_y)/\text{B}(\boldsymbol{\alpha})\}}, \end{aligned} \tag{13}$$

with $n_{hy} = \sum_{i:y_i=y} 1(G_i = h)$, $n_h = \sum_{i=1}^n 1(G_i = h)$, $\bar{\boldsymbol{n}}_y = (n_{1y}, \ldots, n_{Hy})$, $\bar{\boldsymbol{n}} = (n_1, \ldots, n_H)$, $\boldsymbol{\alpha} = (1/H, \ldots, 1/H)$, and $\text{B}(\cdot)$ is the multivariate beta function. It is easy to derive the equalities $\int (\prod_{h=1}^H \upsilon_h^{n_h}) d\Pi_\upsilon = \text{B}(\boldsymbol{\alpha} + \bar{\boldsymbol{n}})/\text{B}(\boldsymbol{\alpha})$ and $\int (\prod_{h=1}^H \upsilon_{hy}^{n_{hy}}) d\Pi_{\upsilon_y} = \text{B}(\boldsymbol{\alpha} + \bar{\boldsymbol{n}}_y)/\text{B}(\boldsymbol{\alpha})$, $y \in \{1,2\}$ exploiting the Dirichlet-multinomial conjugacy.

Although providing a key choice for performing global testing, it is impractical to adopt formulation (12) for each local point null $H_{0l} : \rho_l = 0$ versus $H_{1l} : \rho_l \neq 0$, $l = 1, \ldots, V(V-1)/2$. Hence, we replace local point nulls with small interval nulls $H_{0l} : \rho_l \leq \varepsilon$ versus $H_{1l} : \rho_l > \varepsilon$. This choice allows $\text{pr}[H_{1l} \mid \{\boldsymbol{y}, \mathcal{L}(\boldsymbol{A})\}] = 1 - \text{pr}[H_{0l} \mid \{\boldsymbol{y}, \mathcal{L}(\boldsymbol{A})\}]$ to be easily estimated as the proportion of Gibbs samples in which $\rho_l > \varepsilon$, for each $l = 1, \ldots, V(V-1)/2$. Moreover – as noted in Berger and Sellke (1987) and Berger and Delampady (1987) – testing the small interval hypothesis $H_{0l} : \rho_l \leq \varepsilon$ is in general more realistic and provides – under a Bayesian paradigm – comparable results to those obtained when assessing evidence of $H_{0l} : \rho_l = 0$.

Beside providing key computational benefits, as stated in Proposition 3, our choices induce a prior Π for $p_{\mathcal{Y},\mathcal{L}(\boldsymbol{\mathcal{A}})}$ with full L_1 support over $\mathcal{P}_{2\times|\mathbb{A}_V|}$, meaning that Π can generate a $p_{\mathcal{Y},\mathcal{L}(\boldsymbol{\mathcal{A}})}$ within an arbitrarily small L_1 neighborhood of the true data-generating model $p^0_{\mathcal{Y},\mathcal{L}(\boldsymbol{\mathcal{A}})}$, allowing the truth to fall in a wide class.

Proposition 3. *Based on our priors* $\Pi_y, \Pi_Z, \Pi_X, \Pi_\lambda$, Π_ν, *and letting* $\mathbb{B}_\epsilon(p^0_{\mathcal{Y},\mathcal{L}(\boldsymbol{\mathcal{A}})}) = \{p_{\mathcal{Y},\mathcal{L}(\boldsymbol{\mathcal{A}})} : \sum_{y=1}^2 \sum_{\boldsymbol{a}\in\mathbb{A}_V} |p_{\mathcal{Y},\mathcal{L}(\boldsymbol{\mathcal{A}})}(y,\boldsymbol{a}) - p^0_{\mathcal{Y},\mathcal{L}(\boldsymbol{\mathcal{A}})}(y,\boldsymbol{a})| < \epsilon\}$ *denote the* L_1 *neighborhood around* $p^0_{\mathcal{Y},\mathcal{L}(\boldsymbol{\mathcal{A}})}$, *then for any* $p^0_{\mathcal{Y},\mathcal{L}(\boldsymbol{\mathcal{A}})} \in \mathcal{P}_{2\times|\mathbb{A}_V|}$ *and* $\epsilon > 0$, $\Pi\{\mathbb{B}_\epsilon(p^0_{\mathcal{Y},\mathcal{L}(\boldsymbol{\mathcal{A}})})\} > 0$.

Full prior support is a key property to ensure accurate posterior inference and testing, because without prior support about the true data-generating pmf, the posterior cannot possibly concentrate around the truth. Moreover, as $p_{\mathcal{Y},\mathcal{L}(\boldsymbol{\mathcal{A}})}$ is characterized by finitely many parameters $p_{\mathcal{Y},\mathcal{L}(\boldsymbol{\mathcal{A}})}(y, \boldsymbol{a})$, $y \in \mathbb{Y}$, $\boldsymbol{a} \in \mathbb{A}_V$, Proposition 3 is sufficient to guarantee that the posterior assigns probability one to any arbitrarily small neighborhood of the true joint pmf as $n \to \infty$, meaning that $\Pi[\mathbb{B}_\epsilon(p^0_{\mathcal{Y},\mathcal{L}(\boldsymbol{\mathcal{A}})}) \mid \{y_1, \mathcal{L}(\boldsymbol{A}_1)\}, \ldots, \{y_n, \mathcal{L}(\boldsymbol{A}_n)\}]$ converges almost surely to 1, when the true joint pmf is $p^0_{\mathcal{Y},\mathcal{L}(\boldsymbol{\mathcal{A}})}$.

3.2 Posterior computation

Posterior computation is available via a simple Gibbs sampler, exploiting our representation in Figure 2. Specifically, the Markov Chain Monte Carlo (MCMC) algorithm alternates between the following steps.

1. Sample $p_{\mathcal{Y}}(1) = 1 - p_{\mathcal{Y}}(2)$ from the full conditional $p_{\mathcal{Y}}(1) \mid - \sim \text{Beta}(a + n_1, b + n_2)$, with $n_y = \sum_{i=1}^n 1(y_i = y)$.

2. For each $i = 1, \ldots, n$, update G_i from the discrete variable with probabilities,

$$\text{pr}(G_i = h \mid -) = \frac{\nu_{hy_i} \prod_{l=1}^{V(V-1)/2} (\pi_l^{(h)})^{\mathcal{L}(A_i)_l} (1 - \pi_l^{(h)})^{1-\mathcal{L}(A_i)_l}}{\sum_{q=1}^H \nu_{qy_i} \prod_{l=1}^{V(V-1)/2} (\pi_l^{(q)})^{\mathcal{L}(A_i)_l} (1 - \pi_l^{(q)})^{1-\mathcal{L}(A_i)_l}},$$

   for $h = 1, \ldots, H$, with each $\boldsymbol{\pi}^{(h)}$ factorized as in (9).

3. Given G_i, $i = 1, \ldots, n$, the updating for quantities $\boldsymbol{Z}$, $\boldsymbol{X}^{(h)}$ and $\boldsymbol{\lambda}^{(h)}$, $h = 1, \ldots, H$ proceeds via the recently developed Pólya-gamma data augmentation scheme for Bayesian logistic regression (Polson et al., 2013) as in Durante et al. (2016).

4. Sample the testing indicator T from a Bernoulli with probability (13).

5. If $T = 0$, let $\boldsymbol{\nu}_y = \boldsymbol{\upsilon}$, $y \in \{1, 2\}$ with $\boldsymbol{\upsilon}$ updated from the full conditional Dirichlet $(\upsilon_1, \ldots, \upsilon_H) \mid - \sim \text{Dir}(1/H + n_1, \ldots, 1/H + n_H)$. Otherwise, if $T = 1$, update each $\boldsymbol{\nu}_y$ independently from $(\nu_{1y}, \ldots, \nu_{Hy}) \mid - \sim \text{Dir}(1/H + n_{1y}, \ldots, 1/H + n_{Hy})$.

Since the number of mixture components in (8) and the dimensions of the latent spaces in (9) are not known in practice, we perform posterior computation by fixing H and R at conservative upper bounds. The priors Π_ν and Π_λ are chosen to allow adaptive emptying of the redundant components, with the posteriors for the corresponding parameters controlling unnecessary dimensions concentrated near zero.

4 Simulation studies

We consider simulation studies to evaluate the performance of our method in correctly assessing the global hypothesis of association among the network-valued random variable

$\mathcal{L}(\boldsymbol{\mathcal{A}})$ and the categorical predictor $\mathcal{Y}$, and in identifying local variations in each edge probability across groups.

For comparison we also implement a MANOVA procedure – see e.g. Krzanowski (1988) – to test for global variations across groups in the random vector of summary measures $\boldsymbol{\Theta}$, with realization $\boldsymbol{\theta}_i$ from $\boldsymbol{\Theta}$ comprising the most common network summary statistics – covering network density, transitivity, average path length and assortativity – computed for each simulated network $\boldsymbol{A}_i$. Refer to Kantarci and Labatut (2013) for an overview on these topological network measures, and Bullmore and Sporns (2009), Bullmore and Sporns (2012) for a discussion on their importance in characterizing wiring mechanisms within the brain. For local testing, we compare our procedure to the results obtained when testing on the association between $\mathcal{L}(\mathcal{A})_l$ and $\mathcal{Y}$ via separate two-sided Fisher's exact tests for each $l = 1, \ldots, V(V-1)/2$ – see e.g. Agresti (2002). We consider exact tests to avoid issues arising from χ^2 approximations in sparse tables.

4.1 Simulation settings

We simulate $n = 50$ pairs $(y_i, \boldsymbol{A}_i)$ from our model (5) and (8)–(9), with y_i from a categorical random variable having two equally likely groups $p^0_{\mathcal{Y}}(1) = p^0_{\mathcal{Y}}(2) = 0.5$ and $\boldsymbol{A}_i$ a $V \times V$ network with $V = 20$ nodes. We consider two mixture components, with $\boldsymbol{\pi}^{0(h)}$ defined as in (9). Brain networks are typically characterized by tighter intra-hemispheric than inter-hemispheric connections (Gray Roncal et al., 2013). Hence, we consider two node blocks $\mathbb{V}_1 = \{1, \ldots, 10\}$ and $\mathbb{V}_2 = \{11, \ldots, 20\}$ characterizing left and right hemisphere, respectively, and generate entries in $\boldsymbol{Z}^0$ to favor more likely connections between pairs in the same block, than pairs in different blocks.

To assess performance in local testing, we induce group differences in the connections for a small subset of nodes $\mathbb{V}^* \subset \{1, \ldots, V\}$. To characterize this scenario we let $R^0 = 1$, $\lambda_1^{0(1)} = \lambda_1^{0(2)} = 1$ and consider $X_{v1}^{0(h)} \neq 0$ only for nodes $v \in \mathbb{V}^*$, while fixing the latent coordinates of the remaining nodes to 0. Hence, no variations in edge probabilities are displayed when the mixing probabilities are constant, while only local differences are found when the mixing probabilities shift across groups. Under the dependence scenario, we define group-specific mixing probabilities $\boldsymbol{\nu}_1^0 = (0.8, 0.2)$, $\boldsymbol{\nu}_2^0 = (0.2, 0.8)$. Instead, equal mixing probabilities $\boldsymbol{\nu}_1^0 = \boldsymbol{\nu}_2^0 = (0.5, 0.5)$ are considered under independence. Although we focus on $V = 20$ nodes to facilitate graphical analyses, the mixture representation (8) and the low-rank factorization (9) allows scaling to higher V.

As shown in Figures 3–4, although our dependence simulation scenario may appear – at first – simple, it provides a challenging setting for procedures assessing evidence of global association by testing on changes in the network summary measures. In fact, we choose values $X_{v1}^{0(h)}$ for the nodes $v \in \mathbb{V}^*$ such that the resulting summary statistics for the simulated networks do not display evident variations across groups also in the dependence scenario. Hence, a global test relying on the network summary measures is expected to fail in detecting the association between $\mathcal{Y}$ and $\mathcal{L}(\boldsymbol{\mathcal{A}})$, as the variations in the networks' pmf are mainly local – i.e. in a subset of its marginals $\mathcal{L}(\mathcal{A})_l$. On the other hand, powerful local testing procedures are required to efficiently detect this small set of edge probabilities truly changing across the two groups.

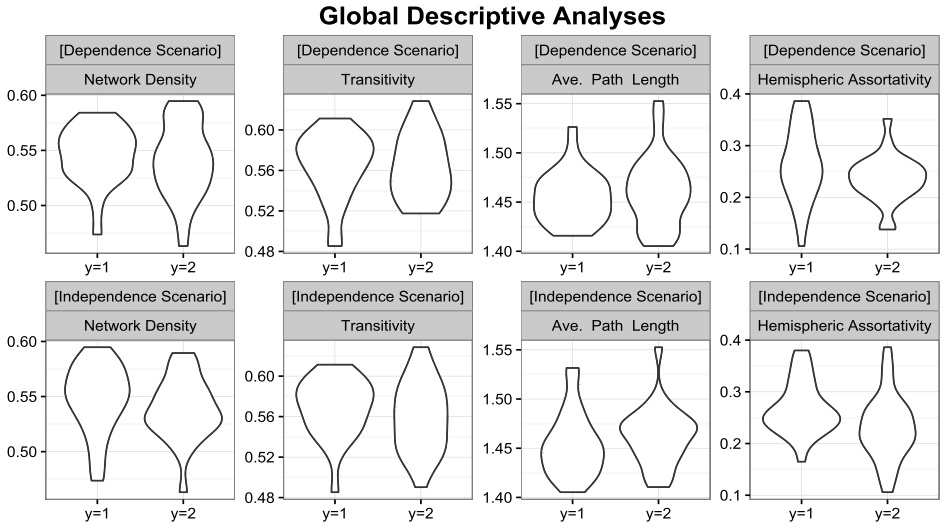

Figure 3: For the two scenarios, observed changes across the two groups for selected network summary statistics. These measures are computed for each simulated network under the two scenarios and summarized via violin plots.

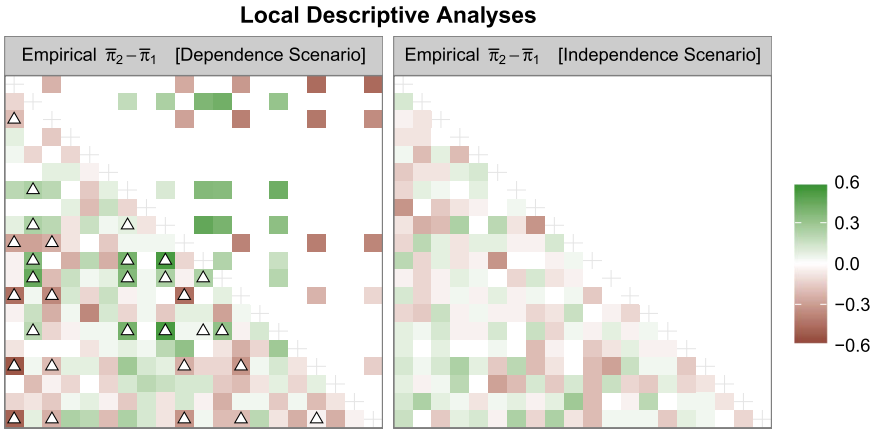

Figure 4: Lower triangular: Group difference between the empirical edge probabilities for each pair of nodes computed from the simulated data. Upper triangular: True group differences from the generative processes considered in the simulations. These quantities are displayed for the dependence (left) and independence (right) scenarios. Triangles highlight edge probabilities which truly differ across groups in the dependence scenario.

In both scenarios, inference is accomplished by considering $H = R = 10$, $\text{pr}(H_1) = \text{pr}(H_0) = 0.5$ and letting $1 - p_{\mathcal{Y}}(2) = p_{\mathcal{Y}}(1) \sim \text{Beta}(1/2, 1/2)$. For priors Π_Z, Π_X and

Π_λ, we choose the same default hyperparameters suggested by Durante et al. (2016). We collect 5,000 Gibbs iterations, discarding the first 1,000. In both scenarios convergence and mixing are assessed via Gelman and Rubin (1992) potential scale reduction factors (PSRF) and effective sample sizes, respectively. The PSRFs are obtained by splitting each chain in four consecutive sub-chains of length 1,000 after burn-in, and comparing between and within sub-chains variances. Convergence and mixing assessments focus on parameters of interest for inference, including the Cramer's V coefficients ρ_l, $l = 1, \ldots, V(V-1)/2$ for local testing and the group-specific edge probability vectors $\bar{\boldsymbol{\pi}}_y$, with elements $\bar{\pi}_{yl} = p_{\mathcal{L}(\mathcal{A})_l|y}(1) = \text{pr}\{\mathcal{L}(\mathcal{A})_l = 1 \mid \mathcal{Y} = y\}$ defined in Proposition 2. This vector coincides with the group-specific mean network structure $\text{E}\{\mathcal{L}(\boldsymbol{\mathcal{A}}) \mid \mathcal{Y} = y\} = \sum_{\boldsymbol{a} \in \mathbb{A}_V} \boldsymbol{a} \times p_{\mathcal{L}(\boldsymbol{\mathcal{A}})|y}(\boldsymbol{a}) = \sum_{h=1}^{H} \nu_{hy} \boldsymbol{\pi}^{(h)}$ under factorization (8). In both scenarios, most of the effective samples sizes are around 2,000 out of 4,000 samples, demonstrating excellent mixing performance. Similarly, all the PSRFs are less than 1.1, providing evidence that convergence has been reached.

4.2 Global and local testing performance

Our testing procedure allows accurate inference on the global association between $\mathcal{L}(\boldsymbol{\mathcal{A}})$ and $\mathcal{Y}$. We obtain $\hat{\text{pr}}[H_1 \mid \{\boldsymbol{y}, \mathcal{L}(\boldsymbol{A})\}] > 0.99$ for the dependence scenario and $\hat{\text{pr}}[H_1 \mid \{\boldsymbol{y}, \mathcal{L}(\boldsymbol{A})\}] < 0.01$ when y_i and $\boldsymbol{A}_i$, $i = 1, \ldots, n$ are generated independently. Instead, the MANOVA testing procedure on the summary statistics vector fails to reject the null hypothesis of no association in both scenarios at a level $\alpha = 0.1$ – as expected. This result further highlights how global network measures may fail in accurately characterizing the whole network architecture.

Focusing on local testing in the dependence scenario, Figure 5 shows how accounting for sparsity and network information – via our dependent mixture of low-rank factor-

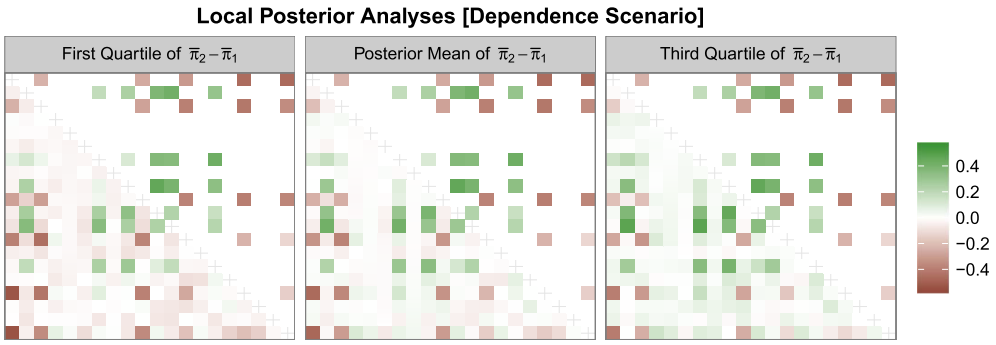

Figure 5: Lower triangular: For the dependence simulation scenario, mean and quartiles of the posterior distribution for the difference between the edge probabilities in the second group $\bar{\pi}_{2l}$ and first group $\bar{\pi}_{1l}$, $l = 1, \ldots, V(V-1)/2$. Upper triangular: For the same scenario, true group difference $\bar{\pi}^0_{2l} - \bar{\pi}^0_{1l}$, $l = 1, \ldots, V(V-1)/2$. In the figure, the pairs of nodes – indexed by l – are re-arranged in matrix form.

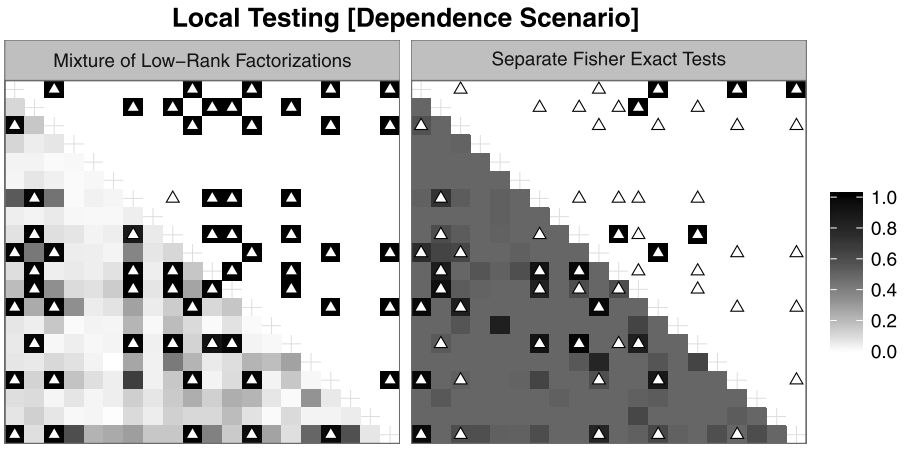

Figure 6: Lower triangular: $\hat{\text{pr}}[H_{1l} \mid \{\boldsymbol{y}, \mathcal{L}(\boldsymbol{A})\}] = \hat{\text{pr}}[\rho_l > 0.1 \mid \{\boldsymbol{y}, \mathcal{L}(\boldsymbol{A})\}]$ (left) and calibrated Fisher's exact tests p-values $1/(1-ep_l \log p_l)$ if $p_l < 1/e$, 0.5 otherwise (right), for each $l = 1, \ldots, V(V-1)/2$. Upper triangular: Rejected local null hypotheses (black), under each method. Triangles highlight edge probabilities which truly differ across groups. In the figure, the pairs of nodes – indexed by l – are re-arranged in matrix form.

izations – provides accurate inference on local variations in edge probabilities, correctly highlighting pairs of nodes whose connectivity differs across groups and explicitly characterizing uncertainty through the posterior distribution. Conducting inference on each pair of nodes separately provides instead poor estimates – refer to left plot in Figure 4 – with the sub-optimality arising from inefficient borrowing of information across the edges. This lack of efficiency strongly affects also the local testing performance as shown in Figure 6, with our procedure having higher power than the one obtained via separate Fisher's exact tests. In Figure 6, each Fisher's exact test p-value is calibrated via $1/(1 - ep_l \log p_l)$ if $p_l < 1/e$ and 0.5 otherwise, to allow better comparison with $\hat{\text{pr}}[H_{1l} \mid \{\boldsymbol{y}, \mathcal{L}(\boldsymbol{A})\}]$ (Sellke et al., 2001). Moreover, we adjust for multiplicity in the Fisher's exact tests by rejecting all the local nulls having a p-value below p^*, with p^* the Benjamini and Hochberg (1995) threshold to maintain a false discovery rate FDR ≤ 0.1. Under our local Bayesian testing procedure we reject all H_{0l} such that $\hat{\text{pr}}[H_{1l} \mid \{\boldsymbol{y}, \mathcal{L}(\boldsymbol{A})\}] > 0.9$, with $\varepsilon = 0.1$. We do not explicitly control for FDR in order to assess whether our Bayesian procedures contain the intrinsic adjustment for multiple testing we expect. According to Figure 6, thresholding the posterior probability of the local alternatives allows implicit adjustment for multiple testings. When explicit FDR control is required, one possibility is to define the threshold following the notion of Bayesian false discovery rate in Newton et al. (2004).

To assess frequentist operating characteristics, we repeated the above simulation exercise for 100 simulated datasets under both dependence and independence scenarios. The MANOVA test is performed under a threshold $\alpha = 0.1$, while the decision rule in the local Fisher's exact tests is based on the Benjamini and Hochberg (1995) threshold

	Type I error	Type II error	FWER	FDR
	Global testing procedure			
Mixture of Low-Rank Factorizations	0.01	0.01		
MANOVA on Summary Measures	0.09	0.90		
	Local testing procedure			
Mixture of Low-Rank Factorizations	0.0004	0.0587	0.0600	0.0023
Separate Fisher's Exact Tests	0.0036	0.5983	0.4000	0.0387

Table 1: Comparison of error rates for our procedure against MANOVA on summary statistics for global testing and separate Fisher's exact tests for local hypotheses.

	Minimum	Mean	Median	Maximum
	Area under the ROC curve (AUC)			
Mixture of Low-Rank Factorizations	0.969	0.999	1.000	1.000
Separate Fisher's Exact Tests	0.810	0.921	0.923	0.989

Table 2: Summary of the AUCs for the 100 simulated datasets in the dependence scenario, to assess performance of local testing at varying thresholds. The ROC curves are defined using the true hypotheses indicators – $\delta_l = 0$ if H_{0l} is true, $\delta_l = 1$ if H_{1l} is true, $l = 1, \ldots, V(V-1)/2$ – and the acceptance or rejection based on our procedure and Fisher's exact tests at varying the thresholds on posterior probabilities and FDR.

to maintain a false discovery rate FDR ≤ 0.1. Under our Bayesian procedure we reject the global null if $\hat{\text{pr}}[H_1 \mid \{\boldsymbol{y}, \mathcal{L}(\boldsymbol{A})\}] > 0.9$. As the prior odds are $\text{pr}(H_1)/\text{pr}(H_0) = 1$, the chosen value 0.9 implies a threshold on the Bayes factor for significance close to the strong evidence bar discussed in Kass and Raftery (1995). According to sensitivity analyses, moderate changes in the threshold do not affect the final conclusions. Consistently with our initial simulation, we reject local nulls if $\hat{\text{pr}}[H_{1l} \mid \{\boldsymbol{y}, \mathcal{L}(\boldsymbol{A})\}] > 0.9$. Also in this case results are not substantially affected by moderate changes in the threshold; hence, we maintain this choice to preserve coherence in our analyses.

Table 1 confirms the superior performance of our approach in maintaining all error rates close to zero, in both global and local testing, while intrinsically adjusting for multiplicity. The information reduction via summary measures for the global test, and the lack of a network structure in the local Fisher's exact tests lead to procedures with substantially less power. Although Table 1 has been constructed using an FDR control of 0.1 in the Fisher's exact tests and a threshold of 0.9 under our local testing procedure, we maintain superior performance allowing the thresholds to vary, as shown in Table 2.

In considering sample size n versus type I and II errors rates, it is interesting to assess the rate at which the posterior probability of the global alternative $\text{pr}[H_1 \mid \{\boldsymbol{y}, \mathcal{L}(\boldsymbol{A})\}]$ converges to 0 and 1 under H_0 and H_1, respectively, as n increases. We evaluate this behavior by simulating 100 datasets as in the previous simulation for increasing sample sizes $n = 20$, $n = 40$ and $n = 100$, and for each scenario. Figure 7 provides histograms showing the estimated posterior probabilities of H_1 for the 100 simulated datasets under the two scenarios and for increasing sample sizes. The separation between scenarios is evident for all sample sizes, with $\hat{\text{pr}}[H_1 \mid \{\boldsymbol{y}, \mathcal{L}(\boldsymbol{A})\}]$ consistently concentrating close to

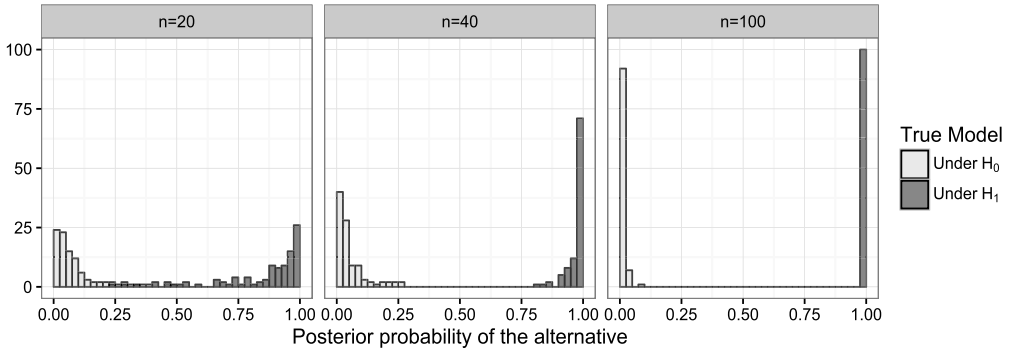

Figure 7: For different n, histograms of the estimated posterior probabilities of the global alternative H_1 for the 100 simulated datasets under dependence and independence.

0 and 1 under the independence and dependence scenario, respectively, as n increases. When $n = 20$ the test has lower power, with 32/100 samples having $\hat{\text{pr}}[H_1 \mid \{\boldsymbol{y}, \mathcal{L}(\boldsymbol{A})\}] < 0.9$ when H_1 is true. However, the type I errors were rare, with 1/100 samples having $\hat{\text{pr}}[H_1 \mid \{\boldsymbol{y}, \mathcal{L}(\boldsymbol{A})\}] > 0.9$ when the data are generated under H_0. These values are very close to 0 when the sample size is increased to $n = 40$ and $n = 100$, with the latter showing strongly concentrated estimates close to 0 and 1, when H_0 is true and H_1 is true, respectively.

4.3 Identifying group differences in more complex functionals

We conclude our simulation studies by considering a scenario in which there is a strong dependence between $\mathcal{L}(\boldsymbol{\mathcal{A}})$ and $\mathcal{Y}$, but this dependence arises from changes in more complex structures, instead of just the edge probabilities. Specifically, we simulate $n = 50$ pairs $(y_i, \boldsymbol{A}_i)$ from our model (5) and (8), with $p^0_{\mathcal{Y}}(1) = p^0_{\mathcal{Y}}(2) = 0.5$ and $\boldsymbol{A}_i$ a $V \times V$ network with $V = 20$ nodes. In defining (8) we consider three components and again split the nodes in two blocks $\mathbb{V}_1 = \{1, \ldots, 10\}$ and $\mathbb{V}_2 = \{11, \ldots, 20\}$, characterizing – for example – the two different hemispheres. When $h = 1$, the vector $\boldsymbol{\pi}^{0(1)}$ characterizes this block structure, with the probability of an edge between pairs of nodes in the same block set at 0.75, while nodes in different blocks have 0.5 probability to be connected. Vectors $\boldsymbol{\pi}^{0(2)}$ and $\boldsymbol{\pi}^{0(3)}$ maintain the same within block probability of 0.75 as in $\boldsymbol{\pi}^{0(1)}$, but have different across block probability. In component $h = 2$ the latter increases by 0.3 – from 0.5 to 0.8 – while in component $h = 3$ this quantity decreases by the same value – from 0.5 to 0.2. As a result, when letting $\boldsymbol{\nu}^0_1 = (1, 0, 0)$ and $\boldsymbol{\nu}^0_2 = (0, 0.5, 0.5)$ it is easy to show that the group-specific edge probabilities – characterizing the distribution of each edge in the two groups – remain equal $\bar{\boldsymbol{\pi}}^0_1 = \bar{\boldsymbol{\pi}}^0_2$, even if the probability mass function jointly assigned to these edges changes across groups $p^0_{\mathcal{L}(\boldsymbol{\mathcal{A}})|1} \neq p^0_{\mathcal{L}(\boldsymbol{\mathcal{A}})|2}$.

This provide a subtle scenario for the several procedures assessing evidence of changes in the brain network across groups, by focusing solely on marginal or expected quantities. These strategies should – correctly – find no difference in edge probabilities, and

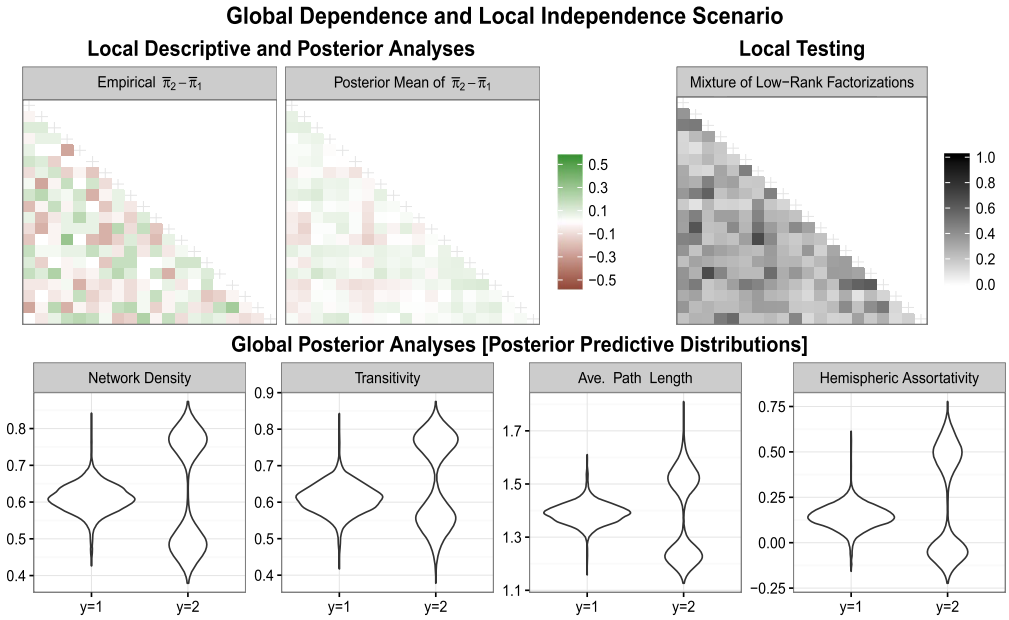

Figure 8: Performance in the final simulation. Upper-left matrix: Group difference between the empirical edge probabilities for each pair of nodes computed from the simulated data (lower triangular) versus the true group difference (upper triangular). Upper-middle matrix: Posterior mean of the difference between the edge probabilities in the two groups (lower triangular) versus true group difference (upper triangular). Upper-right matrix: $\hat{\text{pr}}[H_{1l} \mid \{\boldsymbol{y}, \mathcal{L}(\boldsymbol{A})\}] = \hat{\text{pr}}[\rho_l > 0.1 \mid \{\boldsymbol{y}, \mathcal{L}(\boldsymbol{A})\}]$ (lower triangular) and rejected (black) local null hypotheses (upper triangular), for $l = 1, \ldots, V(V-1)/2$ – rearranged in matrix form. Lower panels: Violin plots representing the posterior predictive distribution of selected network summary statistics in the two groups.

hence may be – wrongly – prone to conclude that the brain network does not change across groups. Underestimating associations may be a dangerous fallacy in understating – for example – the effect of a neurological disorder that induces changes in more complex functionals of the brain network.

We apply our procedures to these simulated data under the same settings of our initial simulations, obtaining very similar effective sample sizes and PSRFs. As shown in the upper panels of Figure 8, the posterior probabilities for all the local alternatives are lower than 0.9, and hence our multiple testing procedure does not reject H_{0l} for every $l = 1, \ldots, V(V-1)/2$. Beside correctly assessing the evidence of no changes in the edge probabilities across the two groups, our global test is able to detect variations in more complex functionals of the brain network. In fact, we obtain $\hat{\text{pr}}[H_1 \mid \{\boldsymbol{y}, \mathcal{L}(\boldsymbol{A})\}] > 0.99$, meaning that, although there is no evidence of changes in edge probabilities across the two groups, the model finds a strong association between $\mathcal{L}(\boldsymbol{\mathcal{A}})$ and $\mathcal{Y}$.

The type of variations in more complex structures can be observed in the lower

panels of Figure 8 showing the posterior predictive distribution of the selected network summary statistics obtained under our statistical model. Although the latter is not analytically available, it is straightforward to simulate from the posterior predictive distribution exploiting our constructive representation in Figure 2 and posterior samples for the quantities in (5) and (8)–(9). Specifically, for each MCMC sample of the parameters in (5) and (8)–(9) – after convergence – we generate a network from our model exploiting the mechanism in Figure 2, to obtain the desired samples from the posterior predictive distribution. According to the lower panels of Figure 8, there are substantial changes in the pmf of the network data across groups. In group one our model infers network summary measures having unimodal distributions, while in the second group we learn substantially different bimodal distributions. This behavior was expected based on our simulation, and hence these results further confirm the accuracy of our global test along with the good performance of our model in flexibly characterizing the distribution of a network-valued random variable and its variations across groups.

5 Application to human brain networks and creativity

We apply our method to the dataset described in the introduction using the same settings as in the simulation examples, but with upper bound H increased to $H = 15$. This choice proves to be sufficient, with components $h = 12, \ldots, 15$ having no observations and redundant dimensions of the latent spaces effectively removed. The efficiency of the Gibbs sampler was very good, with effective sample sizes around 1,500 out of 4,000. Similarly, the PSRFs provide evidence that convergence has been reached, as the highest of these quantities is 1.15. These checks on mixing and convergence are performed for the chains associated with quantities of interest for inference and testing. These include the Cramer's V coefficients ρ_l, $l = 1, \ldots, V(V-1)/2$, the group-specific edge probability vectors $\bar{\boldsymbol{\pi}}_1$, $\bar{\boldsymbol{\pi}}_2$ and the expectation of selected network summary statistics.

Our results provide interesting insights into the global relation between the brain network and creativity, with $\hat{\text{pr}}[H_1 \mid \{\boldsymbol{y}, \mathcal{L}(\boldsymbol{A})\}] = 0.995$ strongly favoring the alternative hypothesis of association between the brain connectivity architecture and the level of creative reasoning. To assess the robustness of our global test, we also performed posterior computation based on datasets that randomly matched the observed group membership variables with a corresponding brain network, effectively removing the possibility of an association. In 10 of these trials we always obtained – as expected – low $\hat{\text{pr}}[H_1 \mid \{\boldsymbol{y}, \mathcal{L}(\boldsymbol{A})\}] \leq 0.2$.

We also attempted to apply the MANOVA test as implemented in the simulation experiments, with the same network statistics – i.e. network density, transitivity, average path length and assortativity by hemisphere. These are popular measures in neuroscience in informing on fundamental properties of the brain network organization, such as small-world, homophily patterns and scale-free behaviors (Bullmore and Sporns, 2009; Rubinov and Sporns, 2010; Bullmore and Sporns, 2012). In our dataset, the average path length was undefined for three subjects, as there were no paths between several pairs of their brain regions. Replacing these undefined shortest path lengths with the maximum

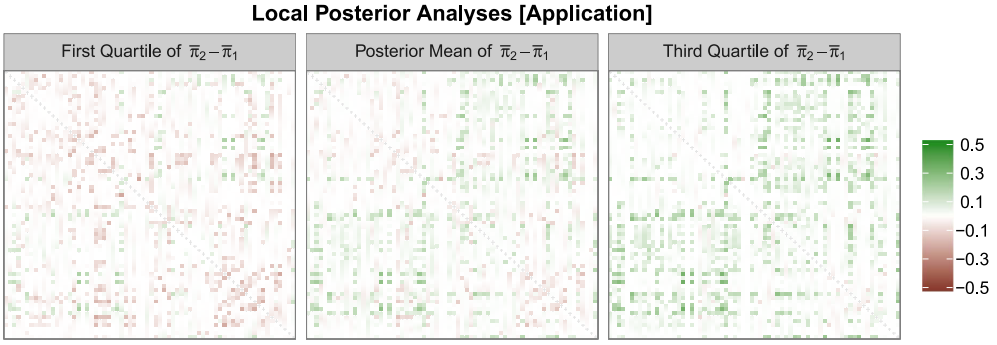

Figure 9: Mean and quartiles of the posterior distribution for the difference $\bar{\pi}_{2l} - \bar{\pi}_{1l}$ between the edge probabilities in high and low creativity groups, for each $l = 1, \ldots, V(V-1)/2$. In the figure, the pairs of brain regions are re-arranged in matrix form.

path length, we observe no significant changes across creativity groups, with a p-value of 0.111. When excluding this topological measure, we obtain a borderline p-value of 0.054. This sensitivity to the choice of summary statistics further motivates tests that avoid choosing topological measures, which is a somewhat arbitrary exercise.

As a secondary focus, we also examined predictive performance of our model. In particular, we considered in-sample edge prediction based on the posterior mean of the edge probabilities in the two groups. This produced excellent results, with an area under the ROC curve (AUC) equal to 0.97. The ROC curve is constructed using the observed edges $\mathcal{L}(A_i)_l$, $i = 1, \ldots, n$, $l = 1, \ldots, V(V-1)/2$ and those predicted with the posterior mean of the group-specific edge probabilities at varying thresholds – using $\hat{\bar{\pi}}_{1l}$ for subjects with $y_i = 1$ and $\hat{\bar{\pi}}_{2l}$ for subjects with $y_i = 2$.

Beside providing a flexible approach for joint modeling of networks and categorical traits, our model also represents a powerful tool to predict y_i given the subject's full brain network structure. In fact, under our formulation, the probability that a subject i has high creativity, conditionally on his brain structural connectivity network $\boldsymbol{A}_i$, is

$$\text{pr}\{\mathcal{Y}_i = 2 \mid \mathcal{L}(\boldsymbol{A}_i)\} = 1 - \text{pr}\{\mathcal{Y}_i = 1 \mid \mathcal{L}(\boldsymbol{A}_i)\} = \frac{p_{\mathcal{Y}}(2) p_{\mathcal{L}(\boldsymbol{\mathcal{A}})|2}(\boldsymbol{a}_i)}{p_{\mathcal{Y}}(2) p_{\mathcal{L}(\boldsymbol{\mathcal{A}})|2}(\boldsymbol{a}_i) + p_{\mathcal{Y}}(1) p_{\mathcal{L}(\boldsymbol{\mathcal{A}})|1}(\boldsymbol{a}_i)},$$

where $\boldsymbol{a}_i = \mathcal{L}(\boldsymbol{A}_i)$ is the network configuration of the ith subject and $p_{\mathcal{L}(\boldsymbol{\mathcal{A}})|y}(\boldsymbol{a}_i)$, $y \in \{1, 2\}$ can be easily computed from (8). We obtain an in-sample AUC = 0.87 in predicting the creativity group y_i using the posterior mean of $\text{pr}\{\mathcal{Y}_i = 2 \mid \mathcal{L}(\boldsymbol{A}_i)\} = 1 - \text{pr}\{\mathcal{Y}_i = 1 \mid \mathcal{L}(\boldsymbol{A}_i)\}$ for each $i = 1, \ldots, n$. Hence, allowing the conditional pmf of the network-valued random variable to shift across groups via group-specific mixing probabilities provides a good characterization of the relation between brain networks and creativity, leading to accurate prediction of the creativity group. Although these results are in-sample, they provide reassurance that the substantial dimensionality reduction underlying our representation does not lead to inadequate fit.

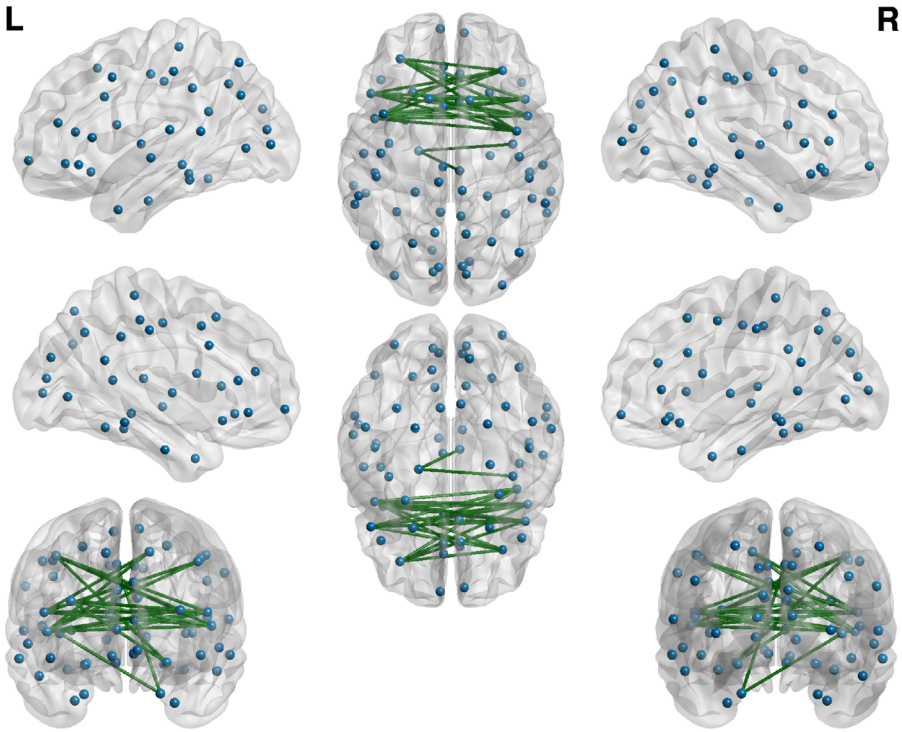

Figure 10: Brain network visualization – from different views – exploiting results from local testing. We only display connections which provide evidence of changes across high and low creativity subjects based on our local tests. Edge color is green – or red – if its estimated probability in high creativity subjects is greater – or smaller – than low creativity ones. Regions' positions are given by their spatial coordinates in the brain.

Figure 9 provides summaries of the posterior distribution for the quantities in $\bar{\boldsymbol{\pi}}_2 - \bar{\boldsymbol{\pi}}_1$, with $\bar{\boldsymbol{\pi}}_2 = \sum_{h=1}^{H} \nu_{h2} \boldsymbol{\pi}^{(h)}$ and $\bar{\boldsymbol{\pi}}_1 = \sum_{h=1}^{H} \nu_{h1} \boldsymbol{\pi}^{(h)}$ encoding the edge probabilities in high and low creativity groups, respectively. Most of these connections have a similar probability in the two groups, with more evident local differences for connections among brain regions in different hemispheres. Highly creative individuals display a higher propensity to form inter-hemispheric connections. Differences in intra-hemispheric circuits are less evident. These findings are confirmed by Figure 10 including also results from our local testing procedure. As in the simulation, we set $\varepsilon = 0.1$ and the decision rule rejects the local null H_{0l} when $\hat{\text{pr}}[H_{1l} \mid \{\boldsymbol{y}, \mathcal{L}(\boldsymbol{A})\}] > 0.9$. These choices provide reasonable settings based on simulations, and results are robust to moderate changes in the thresholds.

Previous studies show that intra-hemispheric connections are more likely than inter-hemispheric connections for healthy individuals (Gray Roncal et al., 2013). This is also evident in our dataset, with subjects having a proportion of intra-hemispheric edges of 0.55 over the total number of possible intra-hemispheric connections, against a propor-

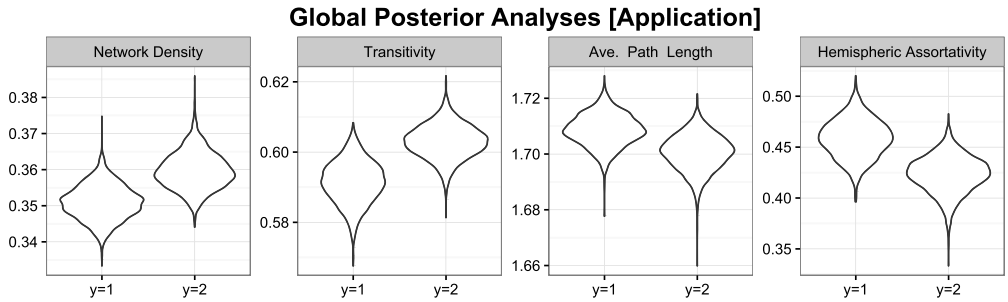

Figure 11: Violin plots representing the posterior distribution for the expectation of selected network summary statistics in the two creativity groups.

tion of about 0.21 for the inter-hemispheric ones. Our estimates in Figure 9 and local tests in Figure 10 highlight differences only in terms of inter-hemispheric connectivity, with highly creative subjects having a stronger propensity to connect regions in different hemispheres. This is consistent with the idea that creative innovations arise from communication of brain regions that ordinarily are not connected (Heilman et al., 2003).

These findings contribute to the ongoing debate on the sources of creativity in the human brain, with original theories considering the right-hemisphere as the seat of creative thinking, and more recent empirical analyses highlighting the importance of the level of communication between the two hemispheres of the brain; see Sawyer (2012), Shobe et al. (2009) and the references cited therein. Beside the differences in techniques to monitor brain networks and measure creativity, as stated in Arden et al. (2010), previous lack of agreement is likely due to the absence of a unifying approach to statistical inference in this field. Our method addresses this issue, while essentially supporting modern theories considering creativity as a result of cooperating hemispheres.

According to Figure 10, the differences in terms of inter-hemispheric connectivity are found mainly in the frontal lobe, where the co-activation circuits in the high creativity group are denser. This is in line with recent findings highlighting the major role of the frontal lobe in creative cognition (Carlsson et al., 2000; Jung et al., 2010; Takeuchi et al., 2010). Previous analyses focus on variations in the activity of each region in isolation, with Carlsson et al. (2000) and Takeuchi et al. (2010) noticing an increase in cerebral blood flow and fractional anisotropy, respectively, for highly creative subjects, and Jung et al. (2010) showing a negative association between creativity and cortical thickness in frontal regions. We instead provide inference on the interconnections among these regions, with increased bilateral frontal connectivity for highly creative subjects, consistent with both the attempt to enhance frontal activity as suggested by Carlsson et al. (2000) and Takeuchi et al. (2010) or reduce it according to Jung et al. (2010).

Figure 11 shows the effect of the increased inter-hemispheric frontal connectivity – in high creativity subjects – on the posterior distribution of the key expected network summary statistics in the two groups. Although the expectation for most of these quantities cannot be analytically derived as a function of the parameters in (8)–(9), it is

straightforward to obtain posterior samples for the previous measures via Monte Carlo methods exploiting the constructive representation in Figure 2. According to Figure 11, the brains in high creativity subjects are characterized by an improved architecture – compared to low creativity subjects – with increased connections, higher transitivity and shortest paths connecting pairs of nodes. As expected, also hemispheric assortativity decreases. This is consistent with our local testing procedure providing evidence of increased inter-hemispheric activity and unchanged intra-hemispheric connectivity structures across the two groups. Previous results are also indicative of small-world structures in highlighting high transitivity and low average path length, with brains for high creativity subjects having a stronger small-world topology than subjects with low creativity. This is a key property in brain networks (Bullmore and Sporns, 2009).

6 Discussion

This article proposes the first general approach in the literature – to our knowledge – for inference and testing of group differences in network-valued data, without focusing on pre-specified functionals or reducing the network data to summary statistics prior to inference. The creativity application illustrates substantial benefits of our approach in providing a unifying and powerful methodology to perform inferences on group differences in brain networks, in contrast to current practice which applies simple statistical tests based on network summary measures or selected functionals. These tests tend to lack power and be sensitive to the summary statistics and functionals chosen, contributing to the inconsistent results observed in the recent literature. Although we specifically focus on creativity, our method can be applied in many other settings. For example, to infer differences in brain networks with neuropsychiatric diseases. In addition, our approach is applicable to other fields involving network-valued data.

It is interesting to generalize our procedure to the multiple group case with $y_i \in \{1, \ldots, K\}$. This can be accomplished with minor modifications to the two groups case. Specifically, it is sufficient to consider as many mixing probability vectors $\boldsymbol{\nu}_y$ as the total number of groups K, replace the beta prior for $p_{\mathcal{Y}}$ with a Dirichlet, and appropriately modify the Gibbs sampler. Theoretical properties and testing procedures are trivial to extend. Although generalization to the multiple groups case is straightforward, there may be subtleties in capturing ordering in the changes across many groups.

There are other interesting ongoing directions. For example, it is important to allow nonparametric shifts in the pmf associated with the network-valued random variable across non-categorical predictor variables, while developing procedures scaling to a number of nodes much larger than $V = 68$. Focusing on neuroscience applications, another important goal is to develop statistical methods that explicitly take into account errors in constructing the brain connection network, including in alignment and in recovering fiber tracts, taking as input the raw imaging data. Our model partially accounts for these errors via the pmf for the network-valued random variable and the prior distributions for its quantities. However, procedures that explicitly account for this noise, may yield improvements in performance, including better uncertainty quantification.

Finally, it is important to consider generalizations accommodating fiber counts, instead of just binary indicators. Incorporating information on weighted edges, data take the form of multivariate counts, again with network-structured dependence. There are subtleties involved in modeling of multivariate count data. It is common to incorporate latent variables in Poisson factor models (e.g. Dunson and Herring, 2005). Including this generalization requires minor modifications of our current procedures, however, as noted in Canale and Dunson (2011), there is a pitfall in such models due to the dual role of the latent variable component in controlling the degree of dependence and the magnitude of over-dispersion in the marginal distributions. Canale and Dunson (2011) address these issues via a rounded kernel method which improves flexibility in modeling count variables. Our current efforts are aimed at adapting these procedures to develop nonparametric approaches for inference on the distribution of weighted networks.

Supplementary Material

Supplementary Materials for "Bayesian Inference and Testing of Group Differences in Brain Networks" (DOI: 10.1214/16-BA1030SUPP; .pdf). The online supplementary material contains proofs of the Propositions 1, 2 and 3, providing theoretical support for our methodology.

References

Agresti, A. (2002). *Categorical Data Analysis*. Second edition. New York: Wiley. MR1914507. doi: https://doi.org/10.1002/0471249688. 41

Airoldi, E. M., Blei, D. M., Fienberg, S. E., and Xing, E. P. (2008). "Mixed membership stochastic blockmodels." *Journal of Machine Learning Research*, 9: 1981–2014. 35

Arden, R., Chavez, R. S., Grazioplene, R., and Jung, R. E. (2010). "Neuroimaging creativity: A psychometric view." *Behavioural Brain Research*, 214(2): 143–156. doi: https://doi.org/10.1016/j.bbr.2010.05.015. 31, 51

Begg, M. D. and Lagakos, S. (1990). "On the consequences of model misspecification in logistic regression." *Environmental Health Perspectives*, 87: 69–75. 32

Benjamini, Y. and Hochberg, Y. (1995). "Controlling the false discovery rate: A practical and powerful approach to multiple testing." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 57(1): 289–300. MR1325392. 32, 44

Berger, J. O. and Delampady, M. (1987). "Testing precise hypotheses." *Statistical Science*, 2(3): 317–335. MR0920141. doi: https://doi.org/10.1214/ss/1177013238. 39

Berger, J. O. and Sellke, T. (1987). "Testing a point null hypothesis: The irreconcilability of P values and evidence." *Journal of the American Statistical Association*, 82(397): 112–122. MR0883340. doi: https://doi.org/10.1080/01621459.1987.10478397. 39

Bhattacharya, A. and Dunson, D. B. (2011). "Sparse Bayesian infinite factor models." *Biometrika*, 98(2): 291–306. MR2806429. doi: https://doi.org/10.1093/biomet/asr013. 36, 38

Bressler, S. L. and Menon, V. (2010). "Large-scale brain networks in cognition: Emerging methods and principles." *Trends in Cognitive Sciences*, 14(6): 277–290. 29

Bullmore, E. and Sporns, O. (2009). "Complex brain networks: Graph theoretical analysis of structural and functional systems." *Nature Reviews Neuroscience*, 10(3): 186–198. 31, 41, 48, 52

Bullmore, E. and Sporns, O. (2012). "The economy of brain network organization." *Nature Reviews Neuroscience*, 13(5): 336–349. 41, 48

Canale, A. and Dunson, D. B. (2011). "Bayesian kernel mixtures for counts." *Journal of the American Statistical Association*, 106(496): 1528–1539. MR2896854. doi: https://doi.org/10.1198/jasa.2011.tm10552. 53

Carlsson, I., Wendt, P. E., and Risberg, J. (2000). "On the neurobiology of creativity. Differences in frontal activity between high and low creative subjects." *Neuropsychologia*, 38(6): 873–885. 51

Craddock, R. C., Jbabdi, S., Yan, C.-G., Vogelstein, J. T., Castellanos, F. X., Martino, A. D., Kelly, C., Heberlein, K., Colcombe, S., and Milham, M. P. (2013). "Imaging human connectomes at the macroscale." *Nature Methods*, 10(6): 524–539. 29, 30

Daianu, M., Jahanshad, N., Nir, T. M., Toga, A. W., Jack, C. R., Weiner, M. W., and Thompson, P. M. (2013). "Breakdown of brain connectivity between normal aging and Alzheimer's disease: A structural k-core network analysis." *Brain Connectivity*, 3(4): 407–422. doi: https://doi.org/10.1089/brain.2012.0137. 31

Deegan, J. (1976). "The consequences of model misspecification in regression analysis." *Multivariate Behavioral Research*, 11(2): 237–248. doi: https://doi.org/10.1207/s15327906mbr1102_9. 32

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. (2006). "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest." *NeuroImage*, 31(3): 968–980. doi: https://doi.org/10.1016/j.neuroimage.2006.01.021. 30, 32

DiRienzo, A. G. and Lagakos, S. W. (2001). "Effects of model misspecification on tests of no randomized treatment effect arising from Cox's proportional hazards model." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(4): 745–757. MR1872064. doi: https://doi.org/10.1111/1467-9868.00310. 32

Dunson, D. B. and Herring, A. H. (2005). "Bayesian latent variable models for mixed discrete outcomes." *Biostatistics*, 6(1): 11–25. doi: https://doi.org/10.1093/biostatistics/kxh025. 53

Dunson, D. B. and Xing, C. (2009). "Nonparametric Bayes modeling of multivariate categorical data." *Journal of the American Statistical Association*, 104(487): 1042–1051. MR2562004. doi: https://doi.org/10.1198/jasa.2009.tm08439. 37

Durante, D. and Dunson, D. B. (2016). "Supplementary material for "Bayesian Inference and Testing of Group Differences in Brain Networks"." *Bayesian Analysis*. doi: https://doi.org/10.1214/16-BA1030SUPP. 37

Durante, D., Dunson, D. B., and Vogelstein, J. T. (2016). "Nonparametric Bayes modeling of populations of networks." *Journal of the American Statistical Association*, in press. doi: https://doi.org/10.1080/01621459.2016.1219260. 32, 35, 38, 40, 43

Fornito, A., Zalesky, A., and Breakspear, M. (2013). "Graph analysis of the human connectome: Promise, progress, and pitfalls." *NeuroImage*, 80: 426–444. doi: https://doi.org/10.1016/j.neuroimage.2013.04.087. 32

Frank, O. and Strauss, D. (1986). "Markov graphs." *Journal of the American Statistical Association*, 81(395): 832–842. MR0860518. doi: https://doi.org/10.1080/01621459.1986.10478342. 31

Fuster, J. M. (2000). "The Module: Crisis of a paradigm." *Neuron*, 26(1): 51–53. doi: https://doi.org/10.1016/S0896-6273(00)81137-X. 29

Fuster, J. M. (2006). "The cognit: A network model of cortical representation." *International Journal of Psychophysiology*, 60(2): 125–132. doi: https://doi.org/10.1016/j.ijpsycho.2005.12.015. 29

Gelman, A. and Rubin, D. B. (1992). "Inference from iterative simulation using multiple sequences." *Statistical Science*, 7(4): 457–472. doi: https://doi.org/10.1214/ss/1177011136. 43

Gelman, A., Van Dyk, D. A., Huang, Z., and Boscardin, J. W. (2008). "Using redundant parameterizations to fit hierarchical models." *Journal of Computational and Graphical Statistics*, 17(1): 95–122. MR2424797. doi: https://doi.org/10.1198/106186008X287337. 36

Genovese, C. R., Lazar, N. A., and Nichols, T. (2002). "Thresholding of statistical maps in functional neuroimaging using the false discovery rate." *NeuroImage*, 15(4): 870–878. doi: https://doi.org/10.1006/nimg.2001.1037. 29

Ghosh, J. and Dunson, D. B. (2009). "Default prior distributions and efficient posterior computation in Bayesian factor analysis." *Journal of Computational and Graphical Statistics*, 18(2): 306–320. MR2749834. doi: https://doi.org/10.1198/jcgs.2009.07145. 36

Ginestet, C. E., Balanchandran, P., Rosenberg, S., and Kolaczyk, E. D. (2014). "Hypothesis testing for network data in functional neuroimaging." arXiv.1407.5525. 32, 33

Gray Roncal, W., 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. (2013). "MIGRAINE: MRI graph reliability analy-

sis and inference for connectomics." In *IEEE Global Conference on Signal and Information Processing*, 313–316. IEEE. doi: https://doi.org/10.1109/GlobalSIP.2013.6736878. 30, 41, 50

Heilman, K. M., Nadeau, S. E., and Beversdorf, D. O. (2003). "Creative innovation: Possible brain mechanisms." *Neurocase*, 9(5): 369–379. doi: https://doi.org/10.1076/neur.9.5.369.16553. 51

Hoff, P. (2008). "Modeling homophily and stochastic equivalence in symmetric relational data." In *Advances in Neural Information Processing Systems*, 657–664. 35, 36, 38

Hoff, P. D., Raftery, A. E., and Handcock, M. S. (2002). "Latent space approaches to social network analysis." *Journal of the American Statistical Association*, 97(460): 1090–1098. MR1951262. doi: https://doi.org/10.1198/016214502388618906. 35

Hunter, D. R., Goodreau, S. M., and Handcock, M. S. (2008a). "Goodness of fit of social network models." *Journal of the American Statistical Association*, 103(481): 248–258. MR2394635. doi: https://doi.org/10.1198/016214507000000446. 31

Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., and Morris, M. (2008b). "ergm: A package to fit, simulate and diagnose exponential-family models for networks." *Journal of Statistical Software*, 24(3). doi: https://doi.org/10.18637/jss.v024.i03. 31

Jung, R. E., Segall, J. M., Bockholt, H. J., Flores, R. A., Smith, S. M., Chavez, R. S., and Haier, R. J. (2010). "Neuroanatomy of creativity." *Human Brain Mapping*, 31(3): 398–409. doi: https://doi.org/10.1002/hbm.20874. 31, 51

Kantarci, B. and Labatut, V. (2013). "Classification of complex networks based on topological properties." In *2013 International Conference on Cloud and Green Computing*, 297–304. IEEE. doi: https://doi.org/10.1109/CGC.2013.54. 41

Kass, R. E. and Raftery, A. E. (1995). "Bayes factors." *Journal of the American Statistical Association*, 90(430): 773–795. MR3363402. doi: https://doi.org/10.1080/01621459.1995.10476572. 45

Krzanowski, W. (1988). *Principles of Multivariate Analysis: A User's Perspective*. Oxford University Press. MR0969370. 41

Newton, M. A., Noueiry, A., Sarkar, D., and Ahlquist, P. (2004). "Detecting differential gene expression with a semiparametric hierarchical mixture method." *Biostatistics*, 5(2): 155–176. doi: https://doi.org/10.1093/biostatistics/5.2.155. 44

Nowicki, K. and Snijders, T. A. B. (2001). "Estimation and prediction for stochastic blockstructures." *Journal of the American Statistical Association*, 96(455): 1077–1087. MR1947255. doi: https://doi.org/10.1198/016214501753208735. 35

Olde Dubbelink, K. T. E., Hillebrand, A., Stoffers, D., Deijen, J. B., Twisk, J. W. R., Stam, C. J., and Berendse, H. W. (2014). "Disrupted brain network topology in Parkinson's disease: A longitudinal magnetoencephalography study." *Brain*, 137(1): 197–207. doi: https://doi.org/10.1093/brain/awt316. 31

Polson, N. G., Scott, J. G., and Windle, J. (2013). "Bayesian inference for logistic models using Pólya–Gamma latent variables." *Journal of the American Statistical Association*, 108(504): 1339–1349. MR3174712. doi: https://doi.org/10.1080/01621459.2013.829001. 40

Ramsey, J. D., Hanson, S. J., Hanson, C., Halchenko, Y. O., Poldrack, R. A., and Glymour, C. (2010). "Six problems for causal inference from fMRI." *NeuroImage*, 49(2): 1545–1558. doi: https://doi.org/10.1016/j.neuroimage.2009.08.065. 30

Rousseau, J. and Mengersen, K. (2011). "Asymptotic behaviour of the posterior distribution in overfitted mixture models." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(5): 689–710. MR2867454. doi: https://doi.org/10.1111/j.1467-9868.2011.00781.x. 39

Rubinov, M. and Sporns, O. (2010). "Complex network measures of brain connectivity: Uses and interpretations." *NeuroImage*, 52(3): 1059–1069. doi: https://doi.org/10.1016/j.neuroimage.2009.10.003. 31, 48

Sawyer, K. R. (2012). *Explaining Creativity: The Science of Human Innovation*. Oxford University Press. 51

Scott, J. G., Kelly, R. C., Smith, M. A., Zhou, P., and Kass, R. E. (2015). "False discovery rate regression: An application to neural synchrony detection in primary visual cortex." *Journal of the American Statistical Association*, 110(510): 459–471. MR3367240. doi: https://doi.org/10.1080/01621459.2014.990973. 32, 33, 38

Sellke, T., Bayarri, M. J., and Berger, J. O. (2001). "Calibration of p values for testing precise null hypotheses." *The American Statistician*, 55(1): 62–71. MR1818723. doi: https://doi.org/10.1198/000313001300339950. 44

Shobe, E. R., Ross, N. M., and Fleck, J. I. (2009). "Influence of handedness and bilateral eye movements on creativity." *Brain and Cognition*, 71(3): 204–214. doi: https://doi.org/10.1016/j.bandc.2009.08.017. 51

Simpson, S. L., Bowman, F. D., and Laurienti, P. J. (2013). "Analyzing complex functional brain networks: Fusing statistics and network science to understand the brain." *Statistics Surveys*, 7: 1–36. MR3161730. doi: https://doi.org/10.1214/13-SS103. 30

Simpson, S. L., Hayasaka, S., and Laurienti, P. J. (2011). "Exponential random graph modeling for complex brain networks." *PLoS ONE*, 6(5): e20039. doi: https://doi.org/10.1371/journal.pone.0020039. 31

Simpson, S. L., Moussa, M. N., and Laurienti, P. J. (2012). "An exponential random graph modeling approach to creating group-based representative whole-brain connectivity networks." *NeuroImage*, 60(2): 1117–1126. doi: https://doi.org/10.1016/j.neuroimage.2012.01.071. 31

Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., Ramsey, J. D., and Woolrich, M. W. (2011). "Network modelling methods for FMRI." *NeuroImage*, 54(2): 875–891. doi: https://doi.org/10.1016/j.neuroimage.2010.08.063. 30

Sporns, O. (2013). "Structure and function of complex brain networks." *Dialogues in Clinical Neuroscience*, 15(3): 247–262. 30

Stam, C. J. (2014). "Modern network science of neurological disorders." *Nature Reviews Neuroscience*, 15(10): 683–695. 31

Stirling, J. and Elliott, R. (2008). *Introducing Neuropsychology*. Routledge. 29

Takeuchi, H., Taki, Y., Sassa, Y., Hashizume, H., Sekiguchi, A., Fukushima, A., and Kawashima, R. (2010). "White matter structures associated with creativity: Evidence from diffusion tensor imaging." *NeuroImage*, 51(1): 11–18. doi: https://doi.org/10.1016/j.neuroimage.2010.02.035. 51

Tansey, W., Koyejo, O., Poldrack, R. A., and Scott, J. G. (2014). "False discovery rate smoothing." arXiv.1411.6144. 29

Wang, H. and Marron, J. (2007). "Object oriented data analysis: Sets of trees." *The Annals of Statistics*, 35(5): 1849–1873. MR2363955. doi: https://doi.org/10.1214/009053607000000217. 31

Wang, J., He, L., Zheng, H., and Lu, Z.-L. (2014). "Optimizing the magnetization-prepared rapid gradient-echo (MP-RAGE) sequence." *PLoS ONE*, 9(5): e96899. doi: https://doi.org/10.1371/journal.pone.0096899. 29

Wasserman, S. and Pattison, P. (1996). "Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p^*." *Psychometrika*, 61(3): 401–425. MR1424909. doi: https://doi.org/10.1007/BF02294547. 31

Zalesky, A., Fornito, A., and Bullmore, E. T. (2010). "Network-based statistic: Identifying differences in brain networks." *NeuroImage*, 53(4): 1197–1207. doi: https://doi.org/10.1016/j.neuroimage.2010.06.041. 32

Acknowledgments

This work was partially funded by the grant CPDA154381/15 of the University of Padova, Italy, and by the U.S. Office of Naval Research grant N00014-14-1-0245. The authors would like to thank Rex E. Jung and Sephira G. Ryman for the brain connectivity data and creativity scores funded by the John Templeton Foundation grant 22156. The authors are grateful to William Gray Roncal and Joshua T. Vogelstein for help in accessing the data. We also thank the Editor, the Associate Editor and the referees for the valuable comments on a first version of the article.