The incomparable: probabilistic graphs for fMRI data analysis

Abstract
Probabilistic graphs are useful approaches when modelling effective cognitive networks from fMRI data. This paper summarizes three recent developments (Dynamic Bayesian Networks, Gaussian Dynamic Networks and Hierarchical Markov Fields) and discusses their (dis)advantages. We argue that, while these approaches differ, a proper comparison would only be possible based on a synthetic or empirical data set. Without such a benchmark, these techniques are not only incomparable but will also remain unused by applied researchers in the field, as they won't be able to tell which techniques yields the most accurate results for the given data.

I Introduction

Researchers in the social sciences are often using fMRI studies to map behavior onto brain function. The analysis of such "brain-data" is often performed with heavy assumptions of linear regression models. Some studies employ less model based approaches such as independent component analysis. In this paper, we are describing how probabilistic graphs can be used to model fMRI data. We are focusing on Bayesian Networks and Markov Networks. Both are probabilistic graphs that represent the connection (arcs) between variables (nodes). Markov Networks are undirected whereas Bayesian Networks are directed graphs. Furthermore, Markov Networks allow for circular structures, whereas Bayesian Networks do not.
The aim of fMRI data analysis is in most cases to describe which neural patterns are activated and, in some cases, if groups of subjects can be discriminated by these patterns (e.g. control/ experimental group or patients/healthy subjects).

We will first introduce the fMRI data structure, then we will discuss solutions for Bayesian Networks and Markov Networks. In the last part, we will be comparing the two approaches and discuss their limitations.

II fMRI data structure

fMRI data is gathered during a scanning session with a patient in an MRI scanner. It is based on the principle that brain areas that are activated will be using more oxygen than the remaining areas. The BOLD (Blood- oxygen-level dependent) signal is what is mapped onto the resulting data structure. It is important to note, that the BOLD signal is in itself only a measure of how much oxygen is transported in a certain brain area. This does not necessarily tell us at which time point an area was active nor how long or how intense. There is a general difference between resting state fMRI and fMRI scans during which event related reactions of the brain are measured. During a resting state scan, the participant lies in the scanner and is instructed to stay awake but not think about anything in particular. A resting state scan is a sort of baseline

measurement. It has been shown that patients can be categorized into different diagnoses based on such resting state (other term: defaults network) scans. The idea is, that the basic network structure differs between patients and healthy population but also between different patient groups. An event related fMRI scan uses stimuli to elicit activation of certain brain regions, the goal is to map active brain areas to specific stimuli. The main problem of this approach is that the BOLD signal is rather time sensitive. Another problem is the background noise that makes it difficult to map a stimulus reaction to a specific brain area.

When using graphical models to analyse fMRI data, the aim is to show connections between areas of the brain (regional connections). In case of event related fMRI, stimuli information is transmitted through the brain in a specific temporal and regional scheme and connectivity patterns related to these stimuli can be estimated. In resting state fMRI, there is no stimulus information and the focus lies on finding connections in the baseline activity of the network.

III Bayesian Networks in Cognitive Modelling

Bayesian networks are graphical models that employ probabilistic reasoning to represent causal connections. A bayesian network $\boldsymbol{B}$ is defined as a pair $\boldsymbol{B} = (G, P)$, where $G = (V(G), A(G))$ is an acyclic directed graph with a set of vertices, $V(G) = \{X_1, \ldots, X_n\}$, and a set of arcs, $A(G)$. Bayesian networks make use of the decomposition rule, ie the chain rule of conditional probabilities, simplifying the joint probability distribution:

$$P(X_1, X_2,...,X_n) = \prod_{i=1}^{n} P(x_i \mid \pi(X_i))$$

In modelling brain connectivity, the nodes in the bayesian network represent the activated brain regions while the arcs characterize the interactions among the regions. The bayesian network then describes the probability distribution over the activation of brain regions. It as assumed that the nodes follow a Gaussian distribution. These models are referred to as Gaussian BN.

With a neural system of n brain regions, the regions are indexed in a set $I = \{i, i = 1, 2, ...n\}$. The activation of a brain region is measured by the average fMRI time-series over the region with x_i measuring the hemodynamic/BOLD response of region i.

The BN is a graphical structure *s* and a joint distribution over the set of time-series $x = \{x_i : i \in I\}$. Let a_i denote the set of activations of the parents of the region i, then the likelihood of activation can be represented as:

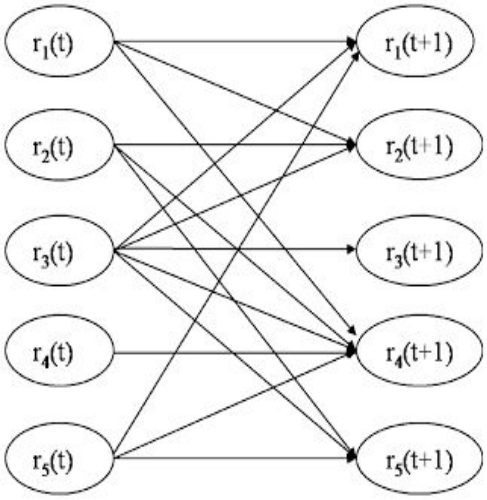
$$P(x \mid \theta) = \prod_{i \in I} P(x_i \mid a_i, \theta_i)$$

Here $\theta = \{\theta_i ; i \in I\}$ representing the parameters of the conditional probabilities.

Dynamic Bayesian networks model the structure of brain connectivity by assuming a class of nonlinear continuous time interactions and modelling the temporal relationships among brain regions (Rajapakse, 2007). They extend the above model to incorporate temporal characteristics of the time-series x; $x(t) = \{x_i(t): i \in I\}$ where the set x(t) represents the activations of n brain regions at time t. To keep the model simple, Rajapakse and Zhou assumed the temporal changes of activations of brain regions to be stationary, first-order Markov chain with transition probabilities independent of t, $P(x(t+1) \mid x(t),...,x(1)) = P(x(t+1) \mid x(t))$. The network represents the connectivity structure between 2 brain scans.

Gaussian BN's assume that the time series of each node, brain region, follows a Gaussian distribution and the BN is a set of linear regression equations. Dynamic BN's, that hold first-order Markov chain and stationary assumptions, model the temporal processes ignored by the Gaussian BN by assuming that the time series of each node follows a multinomial distribution which then discretizes the data into levels.
Discretizing the data in such a way causes a loss of information, which leads to a more recent application by Wu & Wen in which they apply a "Gaussian Dynamic Bayesian Network" method and compare the results with a DBN.

Gaussian dynamic BN
From Rajapaske & Zhou present a DBN study based on 5 brain regions:



Wu & Wen start by considering the above DBN as an "extended Bayesian network" with 10 nodes. Then transformation of the nodes' time series is implemented. They assume there are *T* time point for each variable $r_1,...,r_5$ and their time series in the DBN are $r_1(t),....,r_5(t)$, which correspond to *T-1* time points (for the first 5 variables). Then the last 5 regions: $r_1(t+1),...,r_5(t+1)$ will correspond with the remaining *T-1* time points. In this way the researchers were able to maintain the Gaussian distribution assumption for each of the nodes, and so apply the structure

searching and parameter learning from a Gaussian BN can be applied to the Gaussian dynamic BN.

Structure learning of DBN and an extension to Gaussian DBN

The stationary, first-order Markov chain assumptions of a dynamic BN allow for two layers of n random variables where inter-scan connectivity is always forward. As can be seen from the figure above, this creates 2 columns of nodes and then conditional distributions are only defined for the second layer nodes given the first layer. The DBN is defined as a pair $B = (G, \Theta)$, where G is the transition network structure and Θ represents the set of parameters for the nodes. The arcs are direct dependencies between the connected nodes and missing arcs imply conditional independencies. The joint probability distribution is defined as:

$$P(X_1,...,X_n) = \prod_{i=1}^{n} P(X_i|pa(X_i)) = \prod_{i=1}^{n} \theta_{Xi}|pa(X_i)$$

The parameter $\theta_{Xi}|pa(X_i)$ contains the information of the conditional distribution of the nodes X_i in the 2nd column given its parent nodes $pa(X_i)$ that correspond to the variables in the 1st column. The researchers applied the Bayesian Information Criterion (BIC) based learning approach, a search & score algorithm, to find the network with the highest score out of all possible networks. Search & score algorithms measure the goodness of fit of the BN found and the results are interpreted as the highest score is the better fit. It is a penalized log likelihood score of the data given the network structure. Below is the formula for the BIC, where the first term is the maximized log-likelihood function measuring the degree of fit of the data D given the DBN B. Θ^* the maximum likelihood estimate of the parameters. The second term is the BIC penalty term with d the number of independent parameters and m as the number of data samples.

$$BIC(B|D) \approx \log P(D|B, \Theta^*) - \frac{d}{2} \log m$$

Given the assumptions of the model, node X_i with parent $pa(X_i)$ has the following conditional distribution:

$$p(x_i|pa(x_i)) = \frac{1}{2\pi\sigma i} \exp[-\frac{1}{2\sigma i^2} (x_i-u_i)^2]$$

$$\text{With } u_i = \mu_i + \sum_{x\square \in pa(xi)} b\square(x\square - \mu\square)$$

Where u_i and σi are the conditional mean and conditional variance of node X_i given its parents, b_p is the weight coefficient of the connection from parent node X_p to node X_i. μ_i is the marginal mean of node X_i and $\mu\square$ is the marginal mean of parent node X_p. Then the highest scoring network identified by the BIC approach is seen as a set of multivariate regression equations. The significance of the weight coefficient b_p is tested with stepwise regression.

Furthermore, two indices were created to test the performance of the Gaussian DBN on the synthetic datasets discussed below, they aid in evaluating the learning accuracy. The structure learning index: $S = \frac{N(wrong)}{N(all)}$ and the parameter learning index: $P = \frac{1}{2n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} (c_{ij} - \hat{c}_{ij})^2$. The

structure learning index is the ratio of the number of incorrect connections to the number of all possible connections between the nodes. While P is the squared error between the true connectivity structure c and the estimated structure ĉ.
The researchers applied Rajapakse & Zhou's structure learning method to their discrete DBN for comparison with their Gaussian DBN.

The datasets: Synthetic Data & real fMRI data
We will discuss the most recent research in the Wu & Wen paper. In this paper, the researchers generated a synthetic dataset and compared the results to real fMRI. A vector time series $X_t = (x_{1t}, x_{2t}, \ldots, x_{nt})$ of x brain regions is generated with a first-order multivariate autoregressive model (MAR):

$$X_{t+1} = CX_t + e_t$$

With C is the linear connectivity matrix:

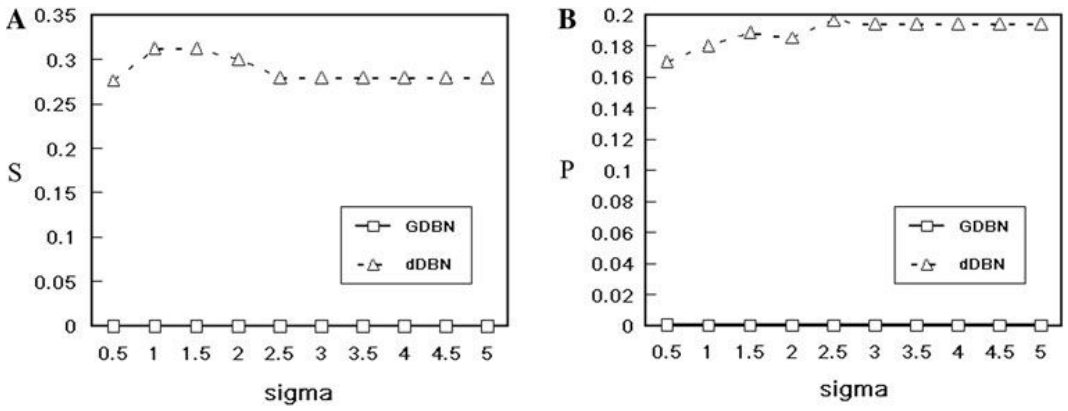
$$C = \begin{bmatrix} -0.9 & 0 & 0.5 & 0 & 0 \\ 0.4 & 0.9 & 0.4 & 0 & 0 \\ 0 & 0 & -0.9 & 0 & 0 \\ -0.5 & 0.8 & -0.9 & -0.9 & 0.5 \\ 0 & -0.6 & -0.7 & 0 & 0.9 \end{bmatrix}$$

And e_t the uncorrelated errors following a Gaussian model with 0 mean and a covariance matrix with diagonal elements equal to σ^2 and off-diagonal elements equal to 0. With white noise parameters to further examine the robustness of the Gaussian DBN inference: the variance of e_t at $\sigma^2 = (0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5)$ and signal to noise (SNR) of uncorrelated Gaussian white noise at levels SNR = (0, 5,10, 15, 20, 25, 30, 35, 40). Finally, 30 sets of data was generated for testing the gaussian DBN in comparison to the discrete DBN.
The real fMRI data is a resting state dataset of 12 Beijing Normal University students, for individual 300 EPI functional volumes were collected. The dataset went through some preprocessing, the first 5 time points were removed and then the remaining were preprocessed for realignment, normalization and smoothing (Wu & wen 2014). A default mode network (DMN) was detected through principle component analysis reduction, independent component analysis separation and back-reconstruction. Time-series of 8 regions of interest (ROI) were chosen based on the DMN; posterior cingulate cortex (PCC), left hippocampus (lHC), right hippocampus (rHC), left inferior parietal cortex (lIPC), right inferior parietal cortex (rIPC), left inferior temporal cortex (lITC), right inferior temporal cortex (rITC) and medial prefrontal cortex (MPFC). The Gaussian DBN estimates the average time courses in each ROI.

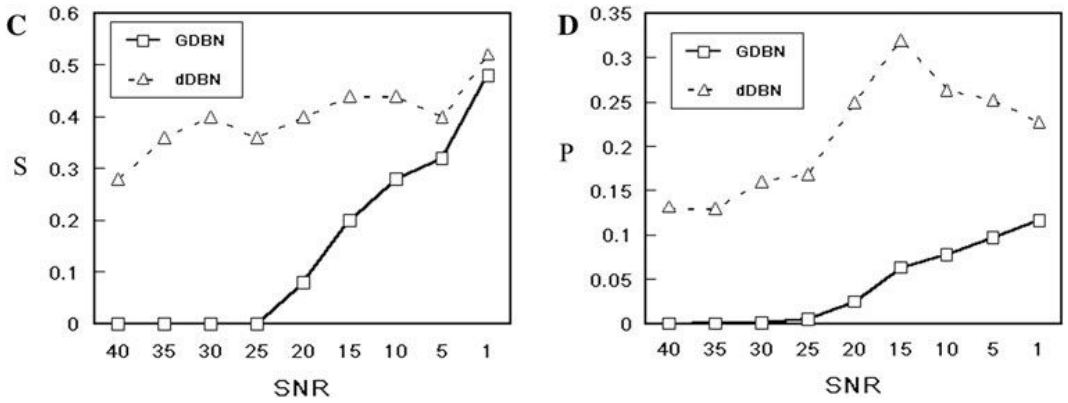
Discussion

1. *Results for structure learning for the synthetic fMRI data*

We discuss the results Wu & Wen obtained from the structure learning for the Gaussian DBN in comparison to the structure learning for discrete DBNs for the generated synthetic dataset. The two plots below: A & B, show the effects of noise variability; e_t with σ^2 on the learned error rate of structure S and parameter index P.



For the GDBN, the learning error rate for both the Structure and Parameter index remained at 0 consistently with the variations in the innovation variance[1], e_t, ie the residual error. These plots indicate that the Gaussian DBN is a more robust method than the discrete DBN.
In the plots below: C & D, the effect of white noise, signal-to-noise, on the learned error rate of structure and parameter index are plotted.

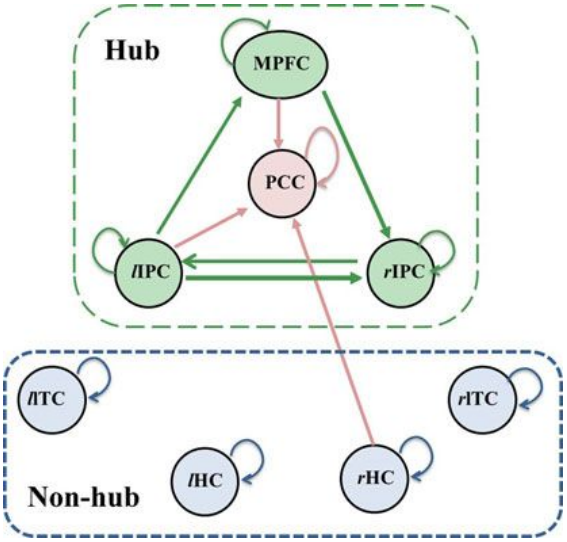


For the structure learning index, S, we can see that the GDBN remained at 0 for high values of SNR, >25, with a steep increase while SNR decreased. While the discrete DBN had higher structure learning index rate over all values for SNR. As for the parameter learning index, it remained very close to 0 with a gradual increase as SNR decreased. Generally, the Gaussian DBN performed better.

2. *Results for structure learning for the real fMRI dataset*

[1] Innovation variance in time-series analysis is the difference between the observed value of a variable at time *t* and the optimal forecast of that value based on information available prior to time *t*.

The figure below shows the effective connectivity structure learned by the Gaussian DBN based on the 8 ROIs detected by the DMN. The first note that comes to mind is how every node is "dynamically influenced by its past activity", which is not strange for brain activity data because it makes sense that an active region will be connected to itself from time point t to time point t+1. We see that there are 5 regions in the Hub component of the network, with only 1 connection to nodes in the non-hub component, from the right hippocampus region to the posterior cingulate cortex. Finally, the PCC appears to be a "confluent node" with no outgoing connections but only receiving information from other regions.



Additionally, two strengths of the Gaussian DBN stand out in comparison to the Gaussian BN. Firstly, the GDBN explicitly takes into consideration the temporal relationships existing in brain regions between consecutive scans. This is not available in BN since firstly, temporal relationships aren't taken into consideration because network structures can have the same skeleton with different arc directions but still arriving at the same marginal likelihood. Secondly, the GDBN is capable of constructing recurrent networks simultaneously maintaining the acyclic property of transition networks. Recurrent networks are important in biological systems, and such a feature is not possible in Gaussian BN, which make Gaussian DBNs a more relevant model for brain connectivity analysis.

Limitations of methods

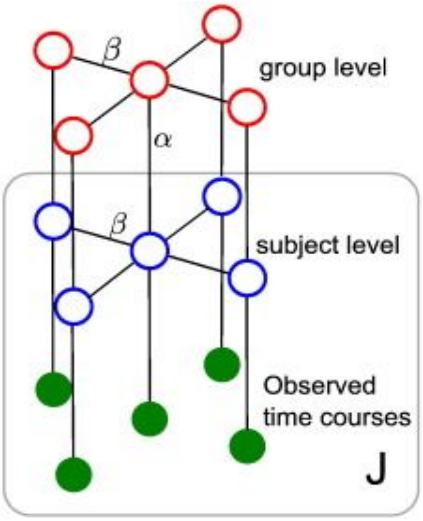
Like all methods, the ones described above bring their limitations. Conventional Gaussian BN ignored the temporal relationships of interactions between brain regions. This limitation, however, was taken care of by Wu & Wen (2014). This was done so by allowing for paths between nodes, thus making conditional dependence between variables feasible. Still, some limitations exist with the application of Gaussian BN. The quality of the algorithm remains unsure when time is extended to higher levels than $t + 1$. Limitations are thus possibly found in concert with brain complexity (e.g. different order modelling, instantaneous interactions between

brain regions, etc.). Gaussian BN furthermore assumes linearity, and it's unlikely that this assumption holds in practice (Wu & Wen, 2014). A big limitation of the Gaussian BN is the feature of acyclicity, which forms a barrier for 'feedback', which is an essential part of biological systems, like the human brain. This limitation brings us to the DBN, where this is easily dealt with. Concerning Dynamic Bayesian Networks (DBN), more limitations exist. Like for Gaussian BN, the problem of instantaneous interactions rises here as well. Time-invariant connectivity between brain regions has been assumed for both methods, while variation has been found by empirical studies (Rajapakse, Zhou, 2007).

IV Markov Fields

In a 2014 paper, Liu, Awate, Anderson and Fletcher approach another problem of fMRI data analysis: modeling group level connectivity patterns from individual scans. Often, researchers would like to describe a group level pattern, to generalize their findings in individuals or to be able to compare different groups. In their paper, they used an image segmentation method that uses information of each subject's scan (individual level) to estimate a group level connectivity pattern. This group level pattern can be seen as the mean of a distribution of connectivity clusters that exist in the population of subjects. Or it other words: a group functional network is shared amongst subjects of one population and the individual networks are just a sample of this group functional network. The goal is to find the most salient characteristics of brain connections of a specific populations.

The researchers are suggesting a hierarchical Markov random field (HMRF) to not only estimate connections on the individual but also on the group level. The HMRF estimates the group-level connectivity map by balancing the individual estimated networks and the group network to include as little variability as possible on the group level and modelling a random factor for individual variability from the data of the subject's scans. In other words, the spatial coherence within a subject (beta parameters in this paper) and the generalization of the individual maps onto a group map (alpha parameter/ pooling factor in this paper) are balanced. The necessary parameters are estimated with the help of Gibbs Sampling, which iteratively estimates the parameters needed to get the posterior distribution.



Schematic representation of the balancing of parameters in HMRF (Liu et al. 2014)

The proposed method is compared to other methods of image segmentation on how high the intersession reliability is and on how much effect perturbations of the data (bootstrapping) have. On the intersession reliability, the proposed HMRF algorithm performs best, with highest overlap between sessions. It must be noted though, that over all subjects, HRMF shows a high variability. It would be interesting if this variability can be found in all subjects or if it is caused by a specific subset. The other methods seem to produce more stable results across subjects in one scanning session (Fig.5, in the original article).
To check the robustness of the method, bootstrapped data (100 samples) are produced and three different methods are run. From these results, seven networks that are also described in the literature are extracted and compared. Unfortunately, the article uses binary maps to show the variance of the bootstrapped results instead of correlations, which makes it difficult to evaluate the effect of the perturbation fully. It seems though that HRMF doesn't miss any important parts in any of the selected networks and has rather low variance between the bootstrapped samples.

Limitations of HMRF
The main downside of the HMRF (or any other pattern recognition method that is used on fMRI data) is the interpretation problem of the results. Once the patterns are found, one has to interpret them in a useful manner. A related problem is to decide how many patterns are kept. In the article by Liu et al., the solution is to use patterns that are closest to patterns described in the literature.
Furthermore, in HMRF, the temporal effect is ignored, as it is in other image segmentation algorithms or algorithms to extract patterns. The result are connected clusters of activation, on the individual and group level, with the idea that group level connectivity can distinguish between different patient groups. Given the temporal structure of fMRI data (and especially the

difference in temporal activation of the physiological measurements), one would expect a successful model to include this dimension.

V Comparison and Future Work

The two presented methods were used either for non-resting or resting state fMRI data. It would be interesting to see if the methods are flexible enough to be used on the other sort of data. For HMRF, one might expect a good result on non-resting state data as circularity can be modelled, but the method will ignore time dependencies. A literature search didn't yield research on this. Bayesian Networks were successfully applied to resting state data to show differences between diseased and healthy subjects (Li et al. 2013). Dynamic BN appear to be more relevant for fMRI data than normal BN's given their inclusion of the temporal characteristic of such datasets. Since they're discrete models, loss of information is too great. The new approach (i.e. Gaussian Dynamic BNs) shows better results for the dynamic BN networks but as we see from the literature this method is relatively new and applications for "disease detection" . Unfortunately, we couldn't source a study comparing DBNs and BNs on resting state data (nor on non-resting state data) to compare both methods.

The big advantage of HMRF is the possibility of estimating group level maps and how individual maps are used as samples from a population map. This thinking might be approached by a form of Hierarchical (D)BNs and might be an interesting topic for future research. Peelen et al. (2009), modelled ICU survival with the help of a hierarchical dynamic BN. Unfortunately, the definition of "hierarchical" is different in their paper though, stating that variables can be expressed as combination of other variables and that there is hence a hierarchy within the nodes of a network. In HMRF, the Bayesian idea of a population distribution (of clusters) stands behind the word "hierarchical".

While researching different approaches of probabilistic graphs for fMRI data, one important flaw was found: there is no consistent benchmarking data used in the articles. Benchmarking might be achieved by synthetic data, but as long as every study uses a slightly different way of simulating, the results of these studies are not comparable (for an overview of problems and benefits of simulated data see: Rodrigues and Andrade, 2015). Alternatively, one could use a sample data set and decide to use such a data set as benchmark. It is important for the entire field of fMRI data analysis research to agree on simulation methods or on a few benchmark data sets. Several universities and research institutes are sharing their fMRI data at websites such as OpenfMRI.
A good example for a benchmarking data set is the denoise data by Kendrick Kay et al.(2013).

VI Conclusion

Proper comparison of these methods on a stimulus as well as a resting state data set would help evaluate these methods and might help applied researchers to choose these data driven methods of analysis over region of interest approaches. For such a comparison, a proper benchmark data set needs to be agreed on.

References:

Kendrick, Ariel, Jonathan, Robert, Brian, 2013, GLMdenoise: a fast, automated technique for denoising task-based fMRI data;
http://journal.frontiersin.org/article/10.3389/fnins.2013.00247/full

Liu, Awate, Anderson, Fletcher, 2014, A functional network estimation method of resting-state fMRI using a hierarchical Markov random field
https://doi-org.ezproxy.leidenuniv.nl:2443/10.1016/j.neuroimage.2014.06.001

Peelen, de Keizer, de Jonge, Bosman, Abu-Hanna, Peek, 2010, Using hierarchical dynamic Bayesian networks to investigate dynamics of organ failure in patients in the Intensive Care Unit
http://www.sciencedirect.com/science/article/pii/S1532046409001373

Rajapakse, Zhou, 2007, Learning effective brain connectivity with dynamic Bayesian networks
http://cs.ru.nl/~peterl/BN/RajapakseZhou2007.pdf

Rodrigues, Andrade, 2015: Synthetic neuronal datasets for benchmarking directed functional connectivity metrics: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4435472/

Wu, Wen, Li, Yoa, 2013, A new dynamic Bayesian network approach for determining effective connectivity from fMRI data
http://rdcu.be/tsw0