NBS
CONNECTOME

Reference Manual for NBS Connectome (v1.2)
December 2012

Contents

Preface

This is the reference manual for NBS Connectome (version 1.2). NBS Connectome is a software package written for Matlab that implements the network-based statistic (NBS) as well as a few other statistical methods for testing hypotheses about the human connectome. Connectomics is a rapidly advancing field, with high profile endeavors such as the NIH's Human Connectome Project capturing considerable scientific and public interest. NBS Connectome is a tool that can be used to make statistical inferences about brain connectivity in health and disease with the imaging data collected as part of such endeavors.

For those that prefer learning by example, NBS Connectome comes preloaded with an example schizophrenia data set. This example data set is loaded by default when the software is opened for the first time.

For those wanting a quick introduction, proceed to Section 3, which contains the bare minimum required to hit the ground running and get started with the software.

Section 1 addresses some basic FAQs about connectomics and the connectome.

The NBS is one of the last steps performed along the typical connectomic data analysis pipeline. In Section 2, we briefly outline some of the tractography and functional connectivity software packages that can be used to get to the NBS stage of this pipeline. We also provide detailed information about the NBS methodology.

Section 4 describes the format of data inputs for NBS Connectome, while advanced topics are covered in Section 5. Finally, examples of the NBS in the scientific literature are provided in an appendix listed at the end of this manual.

We appreciate any feedback, queries and bug reports, or suggestions on how this manual can be improved.

1   FAQs

1.1   What is the NBS?

Basic explanation:

The network-based statistic (NBS; Zalesky et al, 2010) is a validated method for performing statistical analysis on large networks. Numerous studies have used the method to identify connections and networks comprising the human connectome that are associated with an experimental effect or a between-group difference. For example, the NBS has been used to pinpoint faulty brain wiring in psychiatric disorders such a schizophrenia and depression, and to identify enhanced brain connectivity in talented subpopulations such as elite athletes. While the NBS was originally devised for use in connectomics, the method is suited to the analysis of any kind of network where the goal is to independently test the same hypothesis at each and every connection (possibly millions!).

Technical explanation:

The NBS is a nonparametric statistical method to deal with the multiple comparisons problem on a graph. The method is used to control the family-wise error rate (FWER), in the weak sense, when performing mass univariate hypothesis testing on all graph edges. The set of all edges constitutes the family. The NBS is the graph analogue of cluster-based statistical methods used in mass univariate testing on all pixels in an image. Rather than clustering in physical space, the NBS clusters in topological space, where the most basic equivalent of a cluster is a graph component. FWER-corrected p-values are calculated for each component using permutation testing.

1.2   What is the connectome?

The connectome is a buzzword that has grown to encompass basically any network description of whole brain connectivity. The term encompasses all scales, from the micro-scale of individual synaptic connections between neurons to the macro-scale of brain regions and interregional pathways (Sporns et al, 2005). The mapping and analysis of the connectome has defined the new field of connectomics. While the connectome can be mapped with a variety techniques,

magnetic resonance imaging (MRI) is arguably the most pervasive today. With MRI, brain regions represent network nodes and the interconnecting links are traced out using diffusion tractography or functional connectivity techniques. Functional connectivity can yield a vastly different connectome to diffusion tractography, and so speaking of "the connectome" is a misnomer. We conform to the norm and persist in using this misleading term. This has arisen because the connectome originally referred to a comprehensive *structural* description of brain connectivity (Hagmann, 2005; Sporns et al, 2005), but the term has since evolved and broadened to encompass functional connectivity as well. The NBS can be used irrespective of the technique used to map the connectome. Sporns (2012) provides a comprehensive overview of the connectome and connectomics.

1.3   What is NBS Connectome?

NBS Connectome is a software package implementing the NBS as well as a few other methods for testing hypotheses about the human connectome. It is made freely available to the neuroimaging community under the GNU General Public Licence. The software is a collection of Matlab functions that can be easily accessed via graphical user interfaces.

NBS Connectome is platform independent. It can run on any operating system with a modern version of Matlab installed. Matlab is a commercial computing environment developed by MathWorks. Special Matlab toolboxes or functions are not required. Matlab should be available at most academic institutions.

NBS Connectome takes as input a series of connectivity matrices from different groups, or from the same individual during different experimental conditions. The connectivity matrices store numerical values representing the connectivity "strength" between every pair of brain regions. These values are typically derived from neuroimaging data processed with other software packages. The user specifies the hypothesis to be tested at every connection with the general linear model. For example, the user may want to test whether connectivity strength differs between two groups, or whether connectivity strength is correlated with a behavioral measure. The software generates a visualization of the specific connections and networks for which the null hypothesis can be rejected.

1.4   Where can I download NBS Connectome?

The current version (v1.2) can be freely downloaded from:

- NITRC:
  `http://www.nitrc.org/projects/nbs/`

- Brain Connectivity Toolbox (BCT):
  `https://sites.google.com/site/bctnet/comparison/nbs`

The complete package is less than 2MB.

The NBS has also been incorporated into other software packages, including Connectome Viewer (`http://www.connectomics.org`).

1.5 Are there any alternatives to the NBS?

The main alternative to the NBS is the false discovery rate (FDR; Genovese et al, 2002). An option is available in NBS Connectome to use the FDR instead of the NBS. The FDR can be more sensitive to focal effects involving single, isolated connections, whereas the NBS improves power to detect distributed networks spanning multiple connections. However, the story is not that simple. An NBS option is available in the software package to measure the size of an effect using its intensity rather than its extent. The intensity option provides the NBS with greater sensitivity to focal effects.

The NBS and the FDR also differ fundamentally. The FDR enables rejection of the null hypothesis at the level of individual connections, while controlling the false discovery rate; that is, the proportion of false positive connections among all positive connections. The NBS enables rejection of the null hypothesis at the network level (i.e. at the level of a graph component), while controlling the family-wise error rate (FWER); that is, the probability of one or more false positive networks. With the NBS, however, the null hypothesis cannot be rejected for individual connections comprising a significant network. In other words, the FDR trades-off stringent control of false positives for the capacity to perform inference at the level of individual connections. The NBS provides more stringent control of false positives, but at the network level, and exploits any possible interconnected structure inherent to an experimental effect to improve power. We recommend trying both methods.

It is important to remark that if the null hypothesis is true at every connection, both the NBS and the FDR control the FWER at the level of individual connections. That is, if the null hypothesis is true everywhere, both methods will control the chance of a false positive anywhere in the connectome.

A completely different, but complementary approach in connectomics is to analyze topological organization (Rubinov & Sporns, 2010). It has become immensely popular to measure the small-worldness, efficiency, characteristic path length, modularity, etc., of the connectome and then perform statistical inference on these quantities. This approach poses a distinct set of hypotheses compared to performing inference on connectivity strength with the NBS. Network measures express the pattern of interplay between connections at a global level, whereas the NBS is concerned with the strength of connectivity between brain regions at a pairwise level. Depending on the nature of an experimental effect, it may be that only one of two approaches declares the effect to be significant.

Numerous software packages are available to analyze the topological organization of the connectome using network measures:

- Brain Connectivity Toolbox (BCT):
  `https://sites.google.com/site/bctnet/comparison/nbs`

- Matlab Boost Graph Library (BGL):
  `https://github.com/dgleich/matlab-bgl`

- Brainwaver:
  `http://www.nitrc.org/projects/brainwaver`

- Graph Analysis Toolbox (GAT; Hosseini et al, 2012)

1.6   Where can I get further help?

The Matlab functions comprising NBS Connectome were written by Andrew Zalesky. Numerous people contributed to the development, testing and validation of the algorithms implemented in this software package. Please email any of the developers for help:

Andrew Zalesky: `andrewzalesky@gmail.com`
Alex Fornito: `fornitoa@unimelb.edu.au`
Luca Cocchi: `l.cocchi@uq.edu.au`
Ed Bullmore: `etb23@cam.ac.uk`

1.7   Where can I find examples of the NBS?

A selected list of scientific publications featuring use of the NBS is provided as part of the bibliography at the end of this document.

2 NBS Methodology

2.1 Inputs and Outputs

The first step in performing any kind of inference with the NBS is to specify the set of connectivity matrices and the statistical model that will be fitted to each connection. The current version of the software requires the statistical model to be specified in terms of the general linear model (GLM).

Several parameter settings are also provided as input. However, these parameters are not critical to understanding the NBS methodology.

The NBS provides two outputs: the set of connections comprising any network (i.e. graph component) that is found to show a significant effect and a corresponding p-value for each such network.

2.2 General Linear Model

In line with most neuroimaging software packages, such as SPM and FSL, NBS Connectome expects the statistical model to be specified in terms of the GLM. The GLM can handle a diverse range of models, including the t-test, F-test, ANOVA, ANCOVA, multiple linear regression, etc. The GLM is specified in terms of a design matrix and a contrast vector. Statistical modeling and hypothesis testing with the GLM are established subjects that will not be covered here.

The NBS relies on permutation testing to determine significance. The current version of the software uses the method of Freedman & Lane (1983) to perform permutation testing with the GLM. This involves regressing out any nuisance predictors, permuting the resulting residuals and then adding the permuted residuals back to the nuisance signal to give a realization of the data under the null hypothesis. This method recognizes that permuting the raw data is not desirable because it may contain some variability that is explained by the nuisance predictors. In other words, the raw data is not exchangeable under the null hypothesis. It is rather the error terms that can be permuted and they can be estimated under the null hypothesis as the part of the data that is not explained by the nuisance regressors; that is, the residuals (Anderson & Robinson, 2001).

2.3 Connectivity Matrices

The connectivity matrices provided as input the the NBS are typically mapped using other software packages. Mapping connectivity matrices from neuroimaging data is a developing area with many controversies and open questions. For example, the choice of measure used to quantify the connectivity strength between two brain regions; the choice of template used to parcellate the brain into regions representing network nodes; and, the choice of data preprocessing options all remain contentious issues.

Functional MRI:

For resting-state functional MRI data, software packages such as DPARSF (`http://www.nitrc.org/projects/dparsf/`) provide a comprehensive data processing pipeline, extending from the preprocessing of the raw data outputted by the MRI scanner to the mapping of the connectivity matrices that are provided as input to the NBS. Another excellent package for functional connectivity mapping is CONN (`http://www.nitrc.org/projects/conn`). For task-based functional MRI data, the situation is more complex because specialized connectivity measures such as PPI (Friston et al, 1997; O'Reilly, 2012) may be required to disambiguate context-specific connectivity. While PPI was originally intended to map the influence a single seed region exerts on the rest of the brain, this approach has been extended to enable comprehensive mapping of the connectome (Cocchi et al, 2013; Fornito et al, 2012).

Diffusion MRI:

For diffusion MRI data, the number of software packages that perform tractography and tractography visualization are numerous. Some examples include:

- MRtrix:
  `http://www.nitrc.org/projects/mrtrix/`

- TrackVis & Diffusion Toolkit
  `http://www.nitrc.org/projects/trackvis/`

- DSI Studio
  `http://www.nitrc.org/projects/dsistudio/`

- PANDA
  `http://www.nitrc.org/projects/panda/`

These software packages output a set of tractography streamlines that course through the brain, following the trajectories of major white matter fiber pathways.

The task is then to count the total number of streamlines interconnecting each pair of regions. This task is not as trivial as the description here implies and almost all tractography software does not provide options for its implementation. The only exception we know of is PANDA. The user is therefore left to his own devices to develop methods to map connectivity matrices using the streamlines generated by one of the many tractography packages. Hopefully this task will become standardized in the future.

Other data:

Connectivity matrices can also be mapped with EEG/MEG and structural MRI data. However, these modalities are less commonly associated with the connectome. Electron microscopy can also be used to map the connectome at the micro-scale of individual synaptic connections between neurons. But the human connectome is far too expansive for this modality to be used in practice.

It is important to remark that the NBS can be used irrespective of the modality used to map the connectome.

2.4   The NBS Method

The NBS is a simple method that can be explained intuitively without the need for any equations. The first step is to independently test the hypothesis of interest at every connection in the network with an appropriate statistical test. This is known as mass univariate testing. Each connection is therefore endowed with a single test statistic value quantifying the evidence in favor of the null hypothesis.

The second step is to choose a test statistic threshold, sometimes referred to as the primary threshold. The connections with a test statistic value exceeding this threshold are admitted to a set of supra-threshold connections. Connections comprising this set represent potential candidates for which the null hypothesis can be rejected, although at this stage statistical significance cannot be established.

The third step is to identify topological clusters among the set of supra-threshold connections. This step is where the NBS differs from cluster-based statistical methods used in mass univariate testing on all pixels in an image (Nichols & Holmes, 2002). Rather than clustering in physical space, the NBS clusters in topological space, where the most basic equivalent of a cluster is a connected graph component (see Fig. 1). In this context, a connected graph
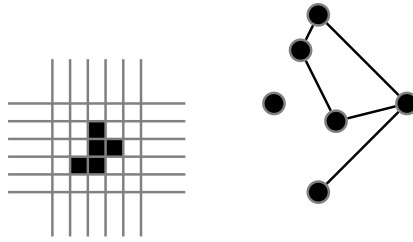
Figure 1: A cluster of pixels in physical space (left) versus a connected graph component in topological space (right). A connected component is the most basic equivalent of a cluster. Both the cluster and the component have an extent of 5.

component is a set of supra-threshold connections for which a path can be found between any two nodes. All connected components can be efficiently identified with a breadth or depth search.

The presence of a component may be evidence of a non-chance structure for which the null hypothesis can be rejected at the level of the structure as a whole, but not for any individual connection alone. The underlying assumption is that the topological configuration of any putative experimental effect is well represented by a component. That is, connections for which the null hypothesis is false are arranged in an interconnected configuration, rather than being confined to a single connection or distributed over several connections that are in isolation of each other. This is a critical assumption inherent to the NBS, and if substantially violated, may render alternative statistical methods favorable.

The final step is to compute a FWER-corrected p-value for each component using permutation testing. The basic premise of permutation testing is that the correspondence between data points and their labels can be randomly rearranged under the null hypothesis without affecting the test statistic. For example, consider a two-sample t-test between samples A and B. If the null hypothesis is true, it does not matter if data points from sample A are randomly exchanged with data points from sample B, although some technical conditions must hold for this to be true (i.e. variance equality). While the t-statistic value is unlikely to be exactly the same when calculated for each random relabeling between samples A and B, any difference only reflects the inherent variance of the data. Of course this would not be the case if the null hypothesis were false.

The data is permuted thousands and thousands of times. Each permutation

involves randomly permuting the data labels according to a permutation vector containing a random permutation of the integers from 1 to the total number of data points. The same permutation vector is used for every connection to preserve any interdependencies between connections. For repeated measures designs, all permutations should be constrained to remain within the same subject (Suckling & Bullmore, 2004). For designs with nuisance regressors, the residuals that remain after regressing out any nuisance predictors are permuted instead of the raw data (see General Linear Model).

For each permutation, the first three steps of the NBS are repeated on the permuted data. In particular, the hypothesis of interest is tested at every connection using the same statistical test, a set of supra-threshold connections is defined using the same threshold and any connected graph components are then identified. The size of the *largest* component is recorded for each permutation, thereby yielding an empirical null distribution for the size of the largest component size.

The one-sided FWER-corrected p-value for a component of given size is then estimated as the proportion of permutations for which the largest component was of the same size or greater.

2.5   How is the size of a component measured?

The simplest measure of component size is the total number of connections it comprises. This is referred to as the component *extent*. Another measure of component size is the sum of test statistic values across all connections comprising the component. This is referred to as the component *intensity*. Extent is suited to detecting an experimental effect that is relatively weak, but extends to encompass many connections. Intensity is suited to a strong, focal effect confined to relatively few connections, although it can also detect strong, distributed effects as well. The concept of extent (area) and intensity (mass) are analogous in cluster-based statistical methods (Bullmore et al, 1999).

Given that the NBS is a nonparametric approach, alternative measures of component size can be defined and trivially incorporated into the method. For example, hybrid measures combining extent and intensity can be used, mirroring developments in cluster-based statistical methods (Hayasaka & Nichols, 2004). Furthermore, while a connected graph component is the simplest definition of a topological cluster, it is not the only definition possible. Spatial pairwise clustering (SPC; Zalesky et al, 2012) is an alternative definition that has recently been incorporated into the NBS.

2.6 Why is the size of the largest component relevant?

Let $\mathcal{H}_0$ denote the case where the null hypothesis is true for every connection. Furthermore, let s_i denote the size of connected graph component i and let α be the nominal FWER. By definition of the FWER, it follows that

$$
\begin{aligned}
\text{FWER} &= \mathbb{P}(\text{ 1 or more components declared significant } |\mathcal{H}_0) \\
&= 1 - \mathbb{P}(\text{ no components declared significant } |\mathcal{H}_0) \\
&= 1 - \mathbb{P}\left(\bigcup_i \{s_i \leq t_\alpha\} | \mathcal{H}_0 \right) \\
&= 1 - \mathbb{P}(\max_i \{s_i\} \leq t_\alpha | \mathcal{H}_0) = \alpha.
\end{aligned}
$$

By storing the size of the largest component, the null distribution $F(t) = \mathbb{P}(\max_i\{s_i\} \leq t | \mathcal{H}_0)$ is estimated empirically. Each permutation contributes a new data point to this empirical estimate. The component size threshold, t_α, for an α-level significance is then determined as $t_\alpha = F^{-1}(1 - \alpha)$.

The crucial detail here is that ensuring the largest component is less than t_α guarantees that all components are less than this critical threshold (Nichols & Hayasaka, 2003).

2.7 How should the test statistic threshold be chosen?

Choosing the test statistic threshold (i.e. primary threshold) is ultimately a matter of experimentation. Fortunately, it is the only parameter associated with the NBS that requires tuning. The arbitrariness of this parameter has been leveled as a criticism of the NBS. The basis of the criticism is that statistical inference with the NBS can be contingent on the choice of test statistic threshold. New networks can appear with a change in threshold, while significant networks can contract/expand to encompass fewer/more connections, or lose their significance all together. A rebuttal to this criticism is that control of the FWER is guaranteed irrespective of the threshold choice. The choice of threshold only affects the sensitivity of the method.

We recommend experimenting with a range of thresholds. Sensitivity to the test statistic threshold may reveal useful information about the nature of the effect. Effects present only at relatively liberal thresholds (e.g. $p < 0.05$, where

we have expressed the threshold as a p-value rather than a test statistic) are likely to be subtle yet topologically extended. Effects present at only conservative thresholds (e.g. $p < 0.001$) are likely to be characterised by strong, topologically focal differences. Effects present across a range of thresholds are likely to be characterised by a combination of both.

3 Getting Started with the Software

3.1 Installation

NBS Connectome is trivial to install. Download and unzip the folder named `NBS1.2.zip`. Open the Matlab window and go to **Set Path**, which is located under the **File** menu. Click on **Add with Subfolders** and specify the path to the unzipped folder. Click **Save**. This step informs Matlab of where the NBS related functions are located.

3.2 Getting Started

Open the Matlab window and change the current folder to the `NBS1.2` directory either with the `cd` command or via the Matlab toolbar. Then type `NBS` at the Matlab command prompt. This will open the NBS window, as shown in Fig 2. Opening the NBS window for the first time populates all fields with the required inputs to exactly replicate the results of the schizophrenia study presented in the originating NBS work (Zalesky et al, 2010).

To get a feel for the software, leave all of the fields unchanged and simply click the **NBS Connectome** button, located at bottom right-hand corner of the window. This prompts the programme to read, verify and process all inputs from the subdirectory named `.../NBS1.2/SchizophreniaExample/`.[1]

Once the **NBS Connectome** button is clicked, a progress bar indicating the proportion of permutations that have been processed appears in the NBS window. The display box located in the bottom left-hand corner of the NBS window is updated in real time with the size of the largest component identified for the current permutation as well as the current best estimate of the most significant p-value. Each new permutation prompts an update.

---

[1]Note that this subdirectory contains the connectivity matrices (`matrices.mat`), design matrix (`designMatrix.txt`), node labels (`nodeLabels.txt`) and node coordinates (`COG.txt`) corresponding to the schizophrenia study (Zalesky et al, 2010). The connectivity matrices were inferred from resting-state functional MRI data. The design matrix specifies a one-sided t-test assessing whether connectivity (Pearson's correlation coefficient) is lower in the schizophrenia sample ($n = 12$) compared to the control group ($n = 15$). The nodes are a subset of 71 brain regions comprising the Automated Anatomical Labeling atlas.

After the requested number of permutations have been generated and processed, the NBSview window opens automatically and displays any significant networks. The NBSview window is shown in Fig. 3. NBSview is a basic utility for viewing brain networks. It is loosely modeled on SPMresults. The NBSview window does not open if no significant networks are found. In the NBSview window, click on a connection or node to display its label. The toolbar buttons in the NBSview window enable the user to change the node diameter, connection width or show all the nodes, rather than just those nodes comprising the significant network.

3.3   Some Things to Try

To get a basic feel for the software, it might be useful to try some of the following trivial experiments. These experiments require all fields in the NBS window to be populated with their default values. To restore default settings, delete the file named `.../NBS1.2/history.mat` and then reopen the NBS window.

- Try changing the test statistic threshold. This is achieved by changing the value of the **Threshold** field in the **Statistical Model** block. Perhaps try a value of 2.5 or 3.5. Click **NBS Connectome** to reevaluate with the new threshold value. For a value of 2.5, the significant network becomes too dense to enable inference, while no significant network should be found for a value of 3.5. Note the reduction in computation time and the absence of the progress bar when reevaluation is performed with a new threshold value. This is because the permutations generated from the first run are stored and reused.

- Try testing for an *increase* in connectivity in the schizophrenia sample. This is achieved by changing the **Contrast** field in the **Statistical Model** block to $[1, -1]$. Click **NBS Connectome** to revaluate with the new contrast. No significant network exhibiting increased connectivity should be found in the schizophrenia sample.

- Try measuring network size with intensity rather than extent. This is achieved by selecting the **Intensity** option from the **Component Size** dropdown menu. Click **NBS Connectome** to reevaluate with the new size measure. The structure of the significant network should remain unchanged when size is measured with intensity, but the associated p-value
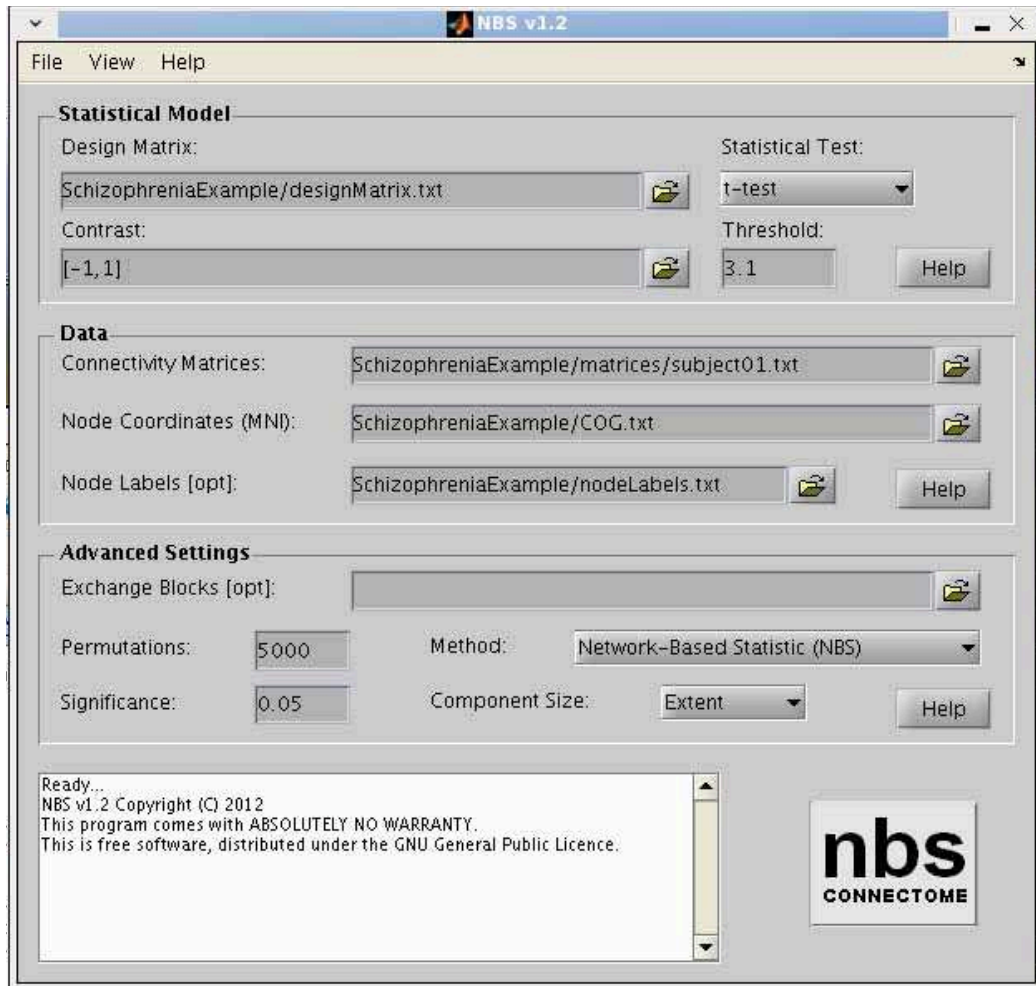
Figure 2: NBS window. The window is divided into three blocks: **Statistical Model**, **Data** and **Advanced Settings**. Opening the NBS window for the first time populates all of the fields with the required inputs to exactly replicate the results presented in the originating NBS work (Zalesky et al, 2010).
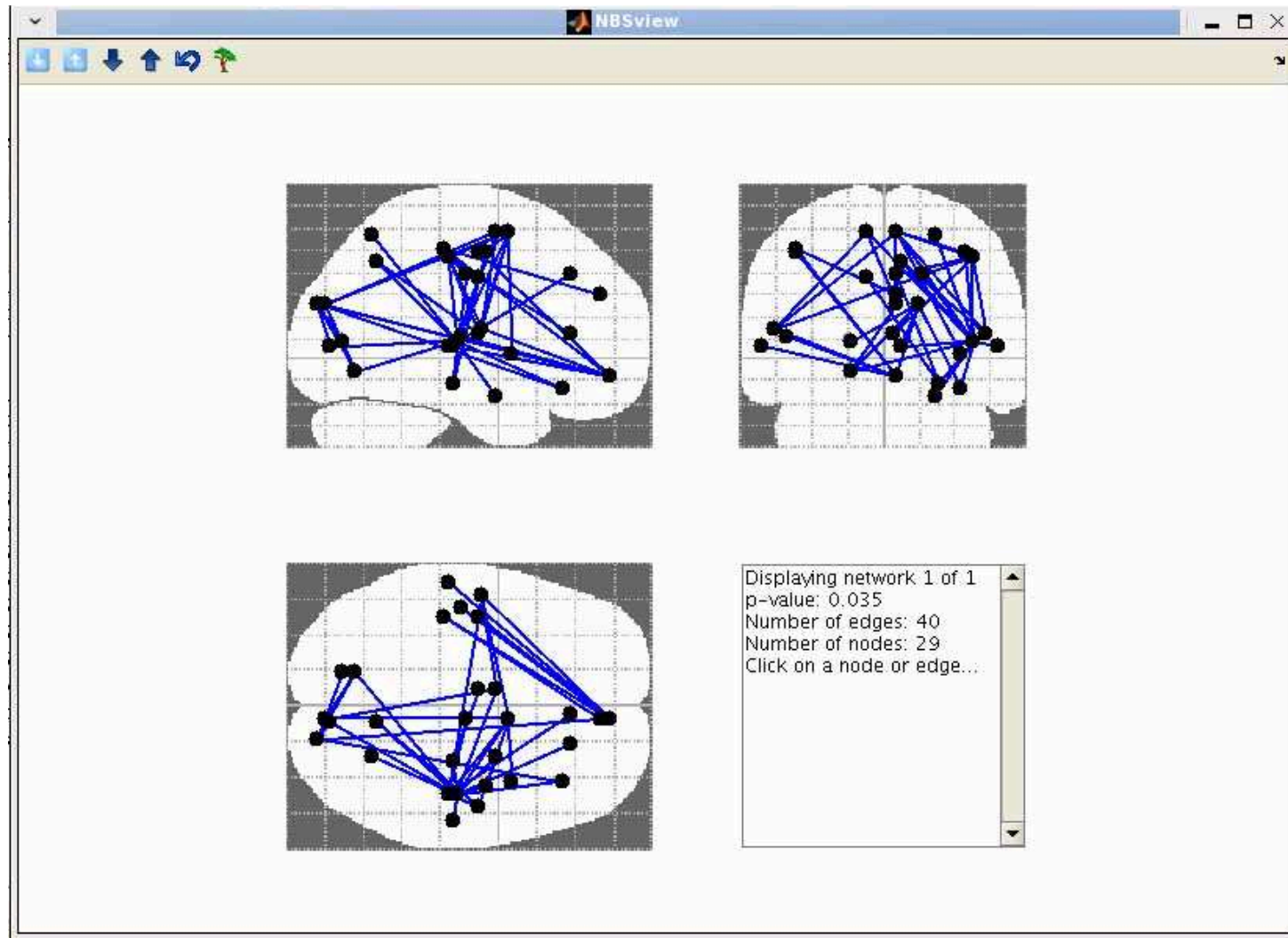
Figure 3: NBSview window. NBSview is a basic utility for viewing brain networks. It is loosely modeled on SPMresults. The network displayed represents reduced resting-state functional connectivity in the schizophrenia sample analyzed in the originating NBS work (Zalesky et al, 2010).

becomes larger compared to when size is measured with extent. This suggests that the effect is distributed and relatively weak.

- Try using the FDR instead of the NBS. This is achieved by selecting the **False Discovery Rate (FDR)** option from the **Method** dropdown menu. It is important to substantially increase the number of permutations generated when using the FDR. To increase the number of permutations, change the value of the **Permutations** field in the **Advanced Settings** block. Perhaps try 50000 permutations. Click **FDR Connectome** to implement the FDR. No significant network should be evident when using the FDR.

3.4   Getting Help

Each of the three blocks comprising the NBS window contains a **Help** button, located in the bottom right-hand corner of the block. Click on the relevant **Help** button to view a detailed explanation of the data format required for each field.

Some users have difficulty in expressing their statistical model and accompanying hypothesis in terms of a design matrix and contrast vector. The help section associated with the **Statistical Model** block contains detailed examples of design matrices and contrast vectors for a t-test, paired t-test (repeated measures), ANOVA and ANCOVA. These examples can be used a templates.

4 Specifying a Statistical Model, Data and Advanced Settings

The NBS window comprises three blocks of inputs: **Statistical Model**, **Data** and **Advanced Settings**. Each block defines a group of related inputs. Inputs are specified with dropdown menus, text files or simply by typing the desired input into the relevant field. Each block of inputs is now described in detail.

4.1 Statistical Model

The **Statistical Model** block contains three fields to specify the GLM and an additional field to specify the test statistic threshold.

Design Matrix:
The design matrix is used to specify the statical model. It can be given either as a:

1. Valid Matlab expression (type the expression)

2. Text file (*.txt)

3. Matlab file (*.mat) containing a single variable

The design matrix should explicitly contain a column of ones (intercept term), if needed. Each column represents an independent variable (predictor) and each row represents a subject or observation. The independent variables can model both tested effects and nuisance covariates. The order in which tested effects and nuisance covariates are added is irrelevant.

The following design matrix performs a t-test with three data points per group, where each column represents a group:

```
Example design matrix: one-sided, two-sample t-test
0 1
0 1
0 1
1 0
1 0
1 0
```

Simply create a text file (*.txt) containing these two columns and specify the file and its path with the browse folder button located next to the **Design Matrix** field.

The help section associated with the **Statistical Model** block contains further examples of design matrices and contrast vectors for a t-test, paired t-test (repeated measures), ANOVA and ANCOVA. Further examples involving the GLM can also be found in the online FEAT user guide:

`http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FEAT/UserGuide`

Contrast:

The contrast vector is used to specify the hypothesis that will be tested and for which a p-value will be estimated. The contrast is given as a valid Matlab expression for a vector that must be the same length as the number of columns in the design matrix.

Given the above design matrix, the following contrast vector specifies a one-sided t-test assessing whether the group modeled by the first column is greater than the group modeled by the second column:

```
Example contrast: group 1 > group 2
[1 -1]
```

Simply type [1 -1] directly into the **Contrast** field. Be sure to include the square brackets.

Statistical Test:

Specify the desired statistical test from the dropdown menu. One of three tests must be selected: **One Sample**, **t-test** or **F-test**. The one sample test randomly flips the sign of each data point for each permutation, which corresponds to a one-sided, one-sample t-test. Note that the one sample test always assesses whether the mean is greater than zero, irrespective of the sign of the contrast vector.

The t-test and the F-test are computed using standard parametric formulae. The t-test is always one-sided.

Threshold:

Specify a value for the test statistic threshold. This value can be fractional. The **Threshold** field disappears if the **False Discovery Rate (FDR)** option is selected from the **Method** dropdown menu. See Section 2.7 for tips on selecting a test statistic threshold.

4.2   Data

The **Data** block contains three fields, of which only the **Connectivity Matrices** field is mandatory.

Connectivity Matrices:

The connectivity matrices can be given either as a:

1. Text file (*.txt), where a separate text file is used to store the connectivity matrix for each subject/observation (provide the path and name of only one of the text files and any others in the same directory will be identified automatically)

2. Valid Matlab expression (type the expression)

3. Matlab file (*.mat) containing a single variable

The Matlab expression or the Matlab file must specify a numeric array with three dimensions, where concatenation across the observations/subjects is in the third dimension. Specifically, element (i, j, k) stores the connectivity value between node i and node j for observation/subject number k. Note that the kth observation/subject is always assumed to correspond to the kth row in the design matrix.

Connectivity matrices must be symmetric about the diagonal. That is, the connectivity strength from node i to node j must be the same as from node j to node i. All connectivity matrices must have the same number of rows/columns and each row/column must correspond to the same brain region across all observations/subjects.

The following is an example of a connectivity matrix comprising 6 brain regions:

```
Example connectivity matrix:
1.0000    0.5997    0.7259    0.6102    0.2896    0.7655
0.5997    1.0000    0.2918    0.5825    0.5418    0.4548
0.7259    0.2918    1.0000    0.7587    0.2357    0.7766
0.6102    0.5825    0.7587    1.0000    0.4586    0.6878
0.2896    0.5418    0.2357    0.4586    1.0000    0.2898
0.7655    0.4548    0.7766    0.6878    0.2898    1.0000
```

Create a separate text file (*.txt) containing the connectivity matrix for each subject/observation and then place all of the files into a common directory. The

files can be named as desired. Specify the path and name of only one of the files with the browse folder button located next to the **Connectivity Matrices** field. All other files in the same directory will be automatically identified and processed.

It is crucial to ensure that each connectivity matrix is assigned to the correct row in the design matrix. If the connectivity matrices are given as separate text files stored in a common directory, the files are first ordered alpha-numerically and the first file in that order is assigned to the first row of the design matrix, etc. For example, the files named `subject01.txt`, `subject02.txt`, `subject03.txt` are assigned to the first, second and third rows of the design matrix, respectively. Note that `Subject11.txt` comes before `Subject1.txt` in alpha-numerical order, hence the importance of including preceding zeros.

Node Coordinates (MNI) [optional]:

Node coordinates are required to display any significant networks. They are specified as a numeric array with three columns (x, y, z) defining MNI coordinates in millimeters. The list of node coordinates must contain as many rows as there are network nodes. It can be given either as a:

1. Valid Matlab expression (type the expression)

2. Text file (*.txt)

3. Matlab file (*.mat) containing a single variable

If node coordinates are not specified, the programme proceeds with executing the prescribed calculations. However, if a significant network is found, it cannot be displayed schematically in NBSview. Only its p-value is shown in the display box located in the bottom left-hand corner of the NBS window.

The following is an example of a coordinate list defining 6 brain regions:

```
Example node coordinate list:
 38.000  -6.000   50.000
-42.000 -10.000   50.000
 18.000  34.000   40.000
-22.000  30.000   44.000
-18.000  46.000  -16.000
 34.000  32.000   34.000
```

Simply create a text file (*.txt) containing these three columns and specify the file and its path with the browse folder button located next to the **Node Coordinates** field.

The first coordinate listed corresponds to the first row/column in the connectivity matrices, etc.

The Matlab script `.../NBS1.2/AAL/getCOG.m` determines the node coordinates (node center of gravity) for an arbitrary node parcellation that is defined in terms of a NIFTI file. Each voxel comprising the same node in the NIFTI file is assumed to have the same intensity. The script will generate a text file listing the coordinates of each node. This list is ordered from lowest to highest intensity. Be sure to set correct values for the voxel dimension variable (`mm`) and the image origin vector (`origin`) in this script.

Node Labels [optional]:

Node labels are an optional input that enable an anatomical name to be assigned to each node. Relevant node labels are displayed in NBSview when a connection or node comprising a significant network is clicked. They do not serve any other purpose. Node labels can be given either as a:

- Valid Matlab expression (type the expression)

- Text file (*.txt), where each line specifies a node label

- Matlab file (*.mat) containing a single variable

The Matlab expression or the Matlab file must specify a cell array of strings, where there are as many strings as there are nodes. Strings must not include whitespace.

The following is an example of a node label list for 6 brain regions:

```
Example node label list:
Precentral_L
Precentral_R
Frontal_Sup_L
Frontal_Sup_R
Frontal_Sup_Orb_R
Frontal_Mid_L
```

Simply create a text file (*.txt) containing these strings and specify the file and its path with the browse folder button located next to the **Node Labels** field.

4.3 Advanced Settings

The **Advanced Settings** block contains options to alter several parameters and settings that are usually best left at their default settings. This block also enables permutations to be constrained for repeated measures (within-subject) designs or interaction effects.

Exchange Blocks [optional]:

Exchange blocks are specified as a numeric vector containing as many rows as there are observations/subjects. Exchange blocks are given either as a:

1. Valid Matlab expression (type the expression)

2. Text file (*.txt), where each line is an observation/subject

3. Matlab file (*.mat) containing a single variable

The following is an example that constrains all permutations to remain within, rows 1 and 4, rows 2 and 5, rows 3 and 6:

```
Example exchange block:
[1;2;3;1;2;3]
```

Simply type [1;2;3;1;2;3] directly into the **Exchange Blocks** field. Be sure to include the square brackets.

Permutations:

Specify the total number of permutations to be generated. The default value is 5000.

Significance:

Specify the FWER corrected significance level (alpha threshold). The default value is 0.05.

Method:

One of three methods to correct for multiple comparisons must be selected from the dropdown menu: **Network-based Statistic (NBS)**, **Spatial Pairwise Clustering (SPC)** or **False Discovery Rate**. Note that SPC is not available as part of the current version.

Component Size [not required for FDR]:

One of two measures must be selected from the dropdown menu to quantify the size of a network (i.e. connected graph component): **Extent** or **Intensity**. See Section 2.5 for further details about measuring network size.

4.4   Generating Results

Once all the input fields have been populated, click on the large square button located in the bottom right-hand corner of the NBS window. The button will be labeled either **NBS Connectome** or **FDR Connectome**, depending on the option selected from the **Method** dropdown menu.

Clicking this button prompts the programme to read, verify and process all inputs. Input fields that were left blank or contain errors become colored red. If all mandatory input fields can be successfully read and verified, the programme proceeds with executing the prescribed set of calculations and no further intervention is required from the user. Otherwise, an error message is shown in the display box located in the bottom left-hand corner of the NBS window. In this case, the input field reported to be in error should be checked. The most common cause of error is an incorrectly formatted design matrix.

A progress bar indicating the proportion of permutations that have been processed appears in the NBS window. The display box is then updated in real time with the size of the largest component identified for the current permutation as well as the current best estimate of the most significant p-value. Each new permutation prompts an update.

The running time is dependent on the network size, the number of permutations and the complexity of the statistical model. The total running time should not exceed a 2-3 minutes in the case of 5000 permutations performed on a network of 90 nodes.

Reevaluation with a different test statistic threshold should be considerably faster (i.e. changing the value in the **Threshold** field and then clicking the large square button again). This is because the permutations from the first run are stored and reused in all subsequent reevaluations. For this reason, the progress bar does not appear when reevaluating with a different test statistic threshold.

If any significant networks are found, the NBSview window opens and displays coronal, axial and sagittal representations of the first significant network. To toggle between multiple significant networks, use the **View** menu in the NBSview

window. The **View** menu does not appear if only one significant network is found. The NBSview window does not open if node coordinates are not specified.

The current results and input fields can be saved with the **Save current** option, which is located under the **File** menu in the NBS window. This creates a Matlab file (*.mat) that can be opened with the **Load previous** option, also located under the **File** menu.

5 Advanced Topics

5.1 How can I access certain values for further analysis?

Subsequent processing and analysis (post-NBS) may require access to the values of certain numerical results. For example, the adjacency matrix of a significant network may be required to generate fancy network visualizations with other software packages. Or, a list of connectivity strengths for each connection comprising a significant network may be required to perform post-hoc testing for correlations with behavioral measures.

The current version of the software stores all numerical results in a Matlab structure called nbs. The nbs structure comprises four substructures: nbs.GLM, nbs.NBS, nbs.STATS and nbs.UI. To access the nbs structure, type global nbs at the Matlab command prompt and then type nbs to see the four substructures:

```
>> global nbs
>> nbs

nbs =

      GLM: [1x1 struct]
      NBS: [1x1 struct]
    STATS: [1x1 struct]
       UI: [1x1 struct]
```

1. nbs.GLM stores the GLM specifications, see documentation in NBSglm.m for details

2. nbs.NBS stores the results, see documentation in NBSview.m for details

3. nbs.STATS stores the test statistic threshold and the test statistics for all permutations, see documentation in NBSstats.m for details

4. nbs.UI stores the user inputs, see documentation in NBSrun.m for details

The data stored in nbs.NBS is usually sufficient for most post-NBS processing and analysis. It can be accessed by typing nbs.NBS at the Matlab command prompt:

```
>> nbs.NBS

ans =

            n: 1
      con_mat: {[74x74 double]}
         pval: 0.0392
    test_stat: [74x74 double]
    node_coor: [74x3 double]
   node_label: {74x1 cell}
```

Note that n is the number of significant networks found and con_mat stores the upper-triangular, binary adjacency matrices for each significant network. One significant network was found in this example.

Adjacency Matrix for Significant Network

To generate a text file (adj.txt) containing a binary adjacency matrix for the first significant network, type the following at the Matlab command prompt:

```
global nbs; adj=nbs.NBS.con_mat{1}+nbs.NBS.con_mat{1}';
dlmwrite('adj.txt',full(adj),'delimiter',' ','precision','%d');
```

Replace {1} with {2}, to repeat for the second significant network, assuming a second network was indeed found.

List of Connections in Significant Network

To print to the screen a list of all connections comprising the first significant network as well as their associated test statistics, type the following at the Matlab command prompt:

```
global nbs; [i,j]=find(nbs.NBS.con_mat{1}); for n=1:length(i)
    i_lab=nbs.NBS.node_label{i(n)};
    j_lab=nbs.NBS.node_label{j(n)};
    stat=nbs.NBS.test_stat(i(n),j(n));
    fprintf('%s to %s. Test stat: %0.2f\n',i_lab,j_lab,stat);
end
```

To print to the screen the list of connectivity strengths for a single connection across all subjects/observations, type the following at the Matlab command prompt:

```
i=1; %Specify node 1 here
j=2; %Specify node 2 here
global nbs; N=nbs.STATS.N; ind_upper=find(triu(ones(N,N),1));
cross_ref=zeros(N,N); cross_ref(ind_upper)=1:length(ind_upper);
cross_ref=cross_ref+cross_ref'; ind=cross_ref(i,j);
fprintf('%0.2f\n',nbs.GLM.y(:,ind));
```

5.2   How can I create network visualizations and figures?

NBSview is a basic utility for viewing brain networks that comes with NBS Connectome. If any significant networks are found, the NBSview window opens and displays coronal, axial and sagittal network representations. To save an image based on the NBSview window, type one of the following commands at the Matlab command prompt:

```
saveas(gcf,'NBSnetwork','jpg'); %Save as jpeg image
saveas(gcf,'NBSnetwork','tif'); %Save as tiff image
saveas(gcf,'NBSnetwork','eps'); %Save as eps image
```

We have often used the PSTricks set of macros for Latex to create network visualizations. However, PSTricks can be difficult to master without an understanding of Latex.

BrainNet Viewer (`http://www.nitrc.org/projects/bnv`) is an excellent brain network visualization tool that has been used to visualize networks identified with the NBS. It is flexible, easy to use and generates three-dimensional network visualizations.

Connectome Viewer (`http://www.connectomics.org`) is another excellent option for brain network visualization (Gerhard et al, 2011). Connectome Viewer comes with a basic NBS plugin enabling all data processing and visualization to be performed within the same software environment. The only limitation is that

the current NBS plugin is limited to a two-sample t-test. A basic understanding of the Python programming language is also required.

5.3   How can certain connections be excluded a priori?

Based on a priori evidence, it may be desirable to exclude certain connections from statistical testing. For example, if it is certain that a white matter pathway does not interconnect a pair of brain regions, it makes little sense to test any hypothesis related to anatomical connectivity for this pair. Doing so would only increase the magnitude of the multiple comparisons problem. However, it is rarely known a priori whether or not a given pair of brain regions is connected, especially in the case of functional connectivity, where indirect anatomical paths can induce direct functional connections.

A common heuristic is to threshold the set of connectivity matrices at the group level and exclude connections that do not exceed the given threshold. For example, a preliminary one-sample t-test can be applied to each connection to assess whether the mean value of connectivity strength differs from zero. Connections that do not differ from zero are then excluded. This assumes zero is an appropriate null connectivity value, which is probably not the case (Jbabdi & Johansen-Berg, 2011). A simpler approach is to compute the group mean connectivity strength for each connection and to exclude connections for which the group mean fails to exceed a given threshold. Criticisms of this approach are the arbitrariness of the threshold choice and the difficulty in dealing with negative connectivity strengths.

With NBS Connectome, connections can be trivially excluded based on a priori evidence. To exclude the connection between brain regions i and j from hypothesis testing, set element (i, j) to zero in every connectivity matrix. In other words, connections with a connectivity strength of zero for all subjects/observations are inherently excluded from hypothesis testing. This means sparse connectivity matrices containing zeros for certain connections across all subjects/observations are inherently excluded.

5.4   How can I use the NBS from the command line?

Refer to the extensive documentation provided in NBSrun.m.

6 Appendices

6.1 Release Log

NBS v1.2 — November 2012

- Fixed the NaN bug in NBSglm that arises when the dependent variable is identically zero for all observations. Force any test statistic value that is NaN to zero. This occurs when the connectivity matrix contains a cell that is identically zero for all observations. Future versions will allow a preliminary test to be performed to exclude from statistical testing any elements in the connectivity matrix that are identically (or sufficiently close to) zero for all observations.

- Fixed the p-value bug in NBSfdr. Previously, the p-value counter was not incremented if the observed test statistic was equal to a randomized test statistic.

- Revised NBSstats to enable fractional component sizes to be shown, which is possible if component size is measured with the intensity option.

NBS v1.1 — August 2012

- Graphical user interface.

- Statistical design specified with general linear model. Covers most statistical designs, including t-test, F-test, ANOVA, ANCOVA, multiple linear regression, etc., and enables modeling of nuisance covariates.

- Exchange blocks for constraining permutations in repeated measures designs.

- NBSview, a basic network viewer modeled on SPMresults.

- Option to measure netowrk size with intensity.

- False discovery rate (FDR) based on nonparametric p-values.

- Permutations stored enabling fast re-computation when experimenting with different test-statistic thresholds.

NBS original version — 2010

- Limited to a two-sample t-test.

6.2 NBS Examples in Scientific Literature: 2011 onwards

- Zalesky A, Fornito A, Bullmore ET. Network-based statistic: identifying differences in brain networks. NeuroImage. 2010. 53(4):1197-1207. *Originating paper

- Alexander-Bloch A, Raznahan A, Bullmore E, Giedd J (2013) The convergence of maturational change and structural covariance in human cortical networks. *J Neuro.* 33(7):2889-2899.

- Bai F, Shu N, Yuan Y, Shi Y, Yu H, Wu D, Wang J, Xia M, He Y, Zhang Z. Topologically convergent and divergent structural connectivity patterns between patients with remitted geriatric depression and amnestic mild cognitive impairment. J Neurosci. 2012. 32(12):4307-4018.

- Cocchi L, Bramati IE, Zalesky A, Furukawa E, Fontenelle LF, Moll J, Tripp G, Mattos P. Altered functional brain connectivity in a non-clinical sample of young adults with attention-deficit/hyperactivity disorder. J Neurosci. 2012. 32(49):17753-17761.

- Cocchi L, Halford GS, Zalesky A, Harding IH, Ramm B, Cutmore T, Shum D, Mattingley JB (2013) Complexity in relational processing predicts changes in functional brain network dynamics *Cereb Cortex.* In press.

- Collin G, de Reus MA, Cahn W, Hulshoff Pol HE, Kahn RS, van den Heuvel MP. Disturbed grey matter coupling in schizophrenia. Eur Neuropsychopharmacol. doi: 10.1016/j.euroneuro.2012.09.001. In press.

- Dwyer D, Fornito A, Harrison BJ, Allen N, Yucel M. Large-scale functional brain networks supporting self-control in adolescence. Poster Presentation: Hum Brain Mapp (OHBM) 2012.

- Fornito A, Yoon J, Zalesky A, Bullmore ET, Carter CS. General and specific functional connectivity disturbances in first episode schizophrenia during cognitive control performance. Biol Psychiatry. 2011. 70(1):64-72.

- Hong S-B, Zalesky A, Cocchi L, Fornito A, Choi E-J, Kim H-H, Suh J-E, Kim C-D, Kim J-W, Yi S-H (2013) Decreased functional brain connectivity in adolescents with internet addiction *PLoS One*. 8(2): e57831.

- Li Y, Jewells V, Kim M, Chen Y, Moon A, Armao D, Troiani L, Markovic-Plese S, Lin W, Shen D. Diffusion tensor imaging based network analysis detects alterations in neuroconnectivity in patients with clinically early relapsing remitting multiple sclerosis. Hum Brain Mapp. In press.

- Li H, Xue Z, Ellmore T, Frye R, Wong S. Network-based analysis reveals stronger local diffusion-based connectivity and different correlations with oral language skills in brains of children with high functioning autism spectrum disorders. Hum Brain Mapp. In press.

- Liu J, Zhao L, Li G, Xiong S, Nan J, Li J, Yuan K, von Deneen KM, Liang F, Qin W, Tian J. Hierarchical alteration of brain structural and functional networks in female migraine sufferers. PLoS One. 2012. 7(12):e51250.

- Shu N, Cao Q, Sun L, An L, Wang P, Xia M, Wang J, Wang Y, He Y. Disrupted Structural Connectome in Attention Deficit Hyperactivity Disorder. Poster Presentation: Hum Brain Mapp (OHBM) 2012.

- Verstraete E, Veldink JH, Mandl RCW, van den Berg LH, van den Heuvel MP. Impaired structural motor connectome in amyotrophic lateral sclerosis. PLoS ONE. 2011. 6:e24239.

- Wang B, Fan Y, Lu M, Li S, Song Z, Peng X, Zhang R, Lin Q, He Y, Wang J, Huang R. Brain anatomical networks in world class gymnasts: A DTI tractography study. Neuroimage. 2012. 65C:476-487.

- Wang J, Xia M, Zhao X, Zhao Z, Zuo X, Dai X, Han Y, Jia J, He Y (2013) Disrupted functional brain connectome in individuals at risk for Alzheimer's disease. *Biol Psych*. 73(5):472-481.

- Zalesky A, Solowij N, Yucel M, Lubman DI, Takagi M, Harding IH, Lorenzetti V, Wang R, Searle K, Pantelis C, Seal M. Effect of long-term cannabis use on axonal fibre connectivity. Brain. 2012. 135(Pt 7):2245-2255.

- Zalesky A, Fornito A, Seal ML, Cocchi L, Westin C-F, Bullmore ET, Egan GF, Pantelis C. Disrupted axonal fiber connectivity in schizophrenia. Biol Psychiatry. 2011. 69(1):80-89.

- Zhang J, Wang J, Wu Q, Kuang W, Huang X, He Y, Gong Q. Disrupted brain connectivity networks in drug-naive, first-episode major depressive disorder. Biol Psychiatry. 2011. 70(4):334-342.

6.3   Last Minute Additions

Some last minute additions that have been brought to our attention:
- Matlab 7.3 (R2006b) or higher is required to avoid an error associated with the Matlab `load` function.

- The NBS command line function `NBSrun.m` has a small memory leak. This should not impact most users in any way whatsoever. However, if the function is called hundreds and hundreds of times, for example, as part of a subroutine in a for-loop, a small amount of memory will be permanently lost to each call. Consequently, if the number of calls to `NBSrun.m` is sufficient, the system may run out of memory. The lost memory can be returned by exiting and restarting Matlab.

References

Anderson MJ & Robinson J (2001) Permutation Tests for Linear Models. Aust NZ J Stat. 43(1):75-88.

Bullmore ET, Suckling J, Overmeyer S, Rabe-Hesketh S, Taylor E, Brammer MJ (1999) Global, voxel, and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain. IEEE Trans Med Imaging. 18(1):32-42.

Cocchi L, Halford GS, Zalesky A, Harding IH, Ramm B, Cutmore T, Shum D, Mattingley JB (2013) Complexity in relational processing predicts changes in functional brain network dynamics. Cereb Cortex. In press.

Fornito A, Harrison BJ, Zalesky A, Simons JS (2012) Competitive and cooperative dynamics of large-scale functional brain networks supporting recollection. Proc Natl Acad Sci U S A. 109(31):12788-12793.

Freedman D, Lane D (1983) A nonstocastic interpretation of reported significance levels. J Bus Econom Statist. 1:292-298.

Friston KJ, Buchel C, Fink GR, Morris J, Rolls E, Dolan R (1997) Psychophysiological and modulatory interactions in Neuroimaging. *NeuroImage* 6:218-229.

Genovese CR, Lazar NA, Nichols TE (2002) Thresholding of statistical maps in functional neuroimaging using the false discovery rate. NeuroImage. 15:870-878.

Gerhard S, Daducci A, Lemkaddem A, Meuli R, Thiran J and Hagmann P (2011) The Connectome Viewer Toolkit: an open source framework to manage, analyze and visualize connectomes. Front Neuroinform. 5:3. doi: 10.3389/fninf.2011.00003

Hagmann P (2005) *From Diffusion MRI to Brain Connectomics*. PhD Thesis, Ecole Polytechnique Fédérale de Lausanne.

Hayasaka S, Nichols TE (2004) Combining voxel intensity and cluster extent with permutation test framework. NeuroImage. 23:54-63.

Hosseini SM, Hoeft F, Kesler SR (2012) GAT: a graph-theoretical analysis toolbox for analyzing between-group differences in large-scale structural and functional brain networks. PLoS ONE. 7(7):e40709.

Jbabdi S, Johansen-Berg H (2011) Tractography: where do we go from here? Brain Connect. 1(3):169-183.

Nichols TE, Hayasaka S (2003) Controlling the familywise rrror rate in functional neuroimaging: a comparative review. Statistical Methods in Medical Research, 12:419-446.

Nichols TE, Holmes AP. Nonparametric permutation tests for functional neuroimaging: a primer with examples (2002) Hum Brain Mapp. 15(1):1-25.

O'Reilly JX, Woolrich MW, Behrens TE, Smith SM, Johansen-Berg H (2012) Tools of the trade: Psychophysiological interactions and functional connectivity. Soc Cogn Affect Neurosci. 7(5):604-609.

Rubinov M, Sporns O (2010) Complex network measures of brain connectivity: uses and interpretations. NeuroImage. 52:1059-1069.

Sporns O, Disocovering the Human Connectome, The MIT Press, 2012.

Sporns O, Tononi G, Kotter R (2005) The human connectome: a structural description of the human brain. PLoS Comput Biol. 1:e42.

Suckling J, Bullmore E (2004) Permutation tests for factorially designed neuroimaging experiments. Hum Brain Mapp. 22(3):193-205.

Zalesky A, Cocchi L, Fornito A, Murray MM, Bullmore E (2012) Connectivity differences in brain networks. NeuroImage. 60(2):1055-1062.

Zalesky A, Fornito A, Bullmore ET (2010) Network-based statistic: Identifying differences in brain networks. NeuroImage. 53(4):1197-1207.