Do different connectivity measures yield similar graph properties from the same data?

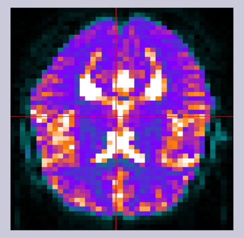
Paul McCarthy Lubica Benuskova
University of Otago, Dunedin, New Zealand

Introduction

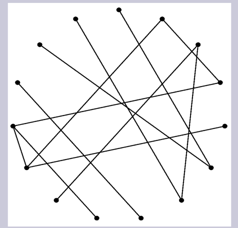
Scientists derive brain functional networks from brain imaging data using different measures of node connectivity. We use three artificial datasets (uniform random, normal/Gaussian random and artificially correlated) and three connectivity measures to look at properties of the resulting unweighted and undirected graphs. We present the results of analysis for the Pearson correlation coefficient, frequency coherence, and a third measure found in the literature, the functional distance ([6]).

Analysing data using graph theory

Graph theory provides some useful techniques for the analysis of brain imaging data, and of time series in general. The degree distribution of a graph gives an indication of the presence of highly connected components in, and highlights the scale-free nature of, the brain. Graph density may be used to compare the levels of interaction within the brains of different subjects. The characteristic path length and clustering coefficient highlight the small-world nature of functional networks in the brain.



$$\Rightarrow \frac{\sum(x_i - \bar{x})\,(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2\,\sum(y_i - \bar{y})^2}} \Rightarrow$$



A graph is created from a data set as follows:

1. Every signal in the data set (e.g. voxels in an fMRI image, or channels in a MEG/EEG recording) is added as a node in the graph.
2. Some measure of association is calculated between every pair of signals in the data set; the association value between a pair of signals determines the existence of an edge between the corresponding nodes in the graph.
3. To generate an unweighted graph, some threshold is applied to the association values; an edge is added to the graph for each value above the threshold.

Association measures

- The *Pearson correlation coefficient* is the standard statistical measure of bivariate linear association, and is widely used in the analysis of functional connectivity in the brain ([5], [9], [10], [3]).
- Similarly, *coherence* is an established measure of association in signal processing, and is useful in assessing whole or frequency-band associations in MEG/EEG data ([4]).
- Shen et. al. ([6]) propose the *functional distance* as a measure of association, the euclidean distance between data points, embedded in a Gaussian kernel.
- There are many more measures of assocation in wide use, including wavelet coherence ([1]) and synchronization likelihood ([7]).

Data

Three artificial data sets were created, each containing 2000 time series of 2000 samples each:

- *Uniform*: randomly generated from a uniform distribution.
- *Normal*: randomly generated from a normal distribution.
- *Cholesky*: randomly generated from a uniform distribution, and artificially correlated via Cholesky factorisation. The correlation matrix contained values uniformly distributed in the range 0.99 – 1.0.

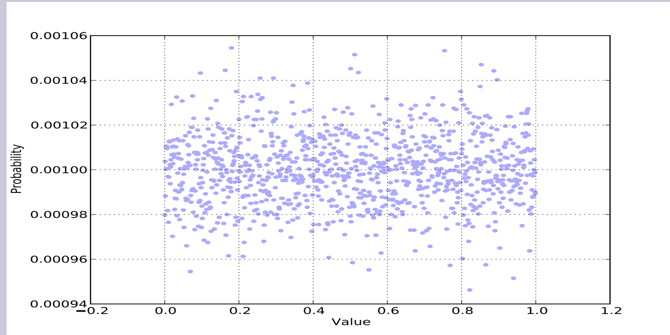
Method

For each of the data sets, the coherence, functional distance, and correlation were calculated to generate 2000x2000 association matrices. Unweighted, undirected graphs were generated from these matrices across a range of thresholds chosen according to the association value distributions.

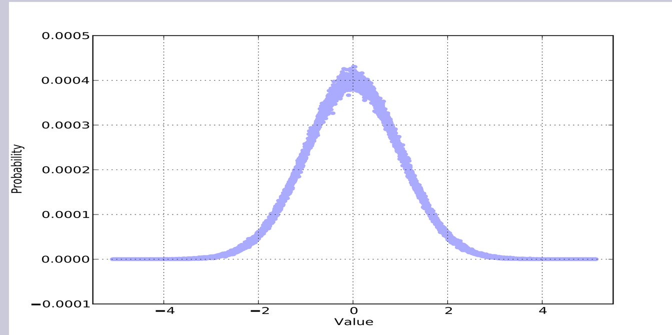
Data distributions

Shown here are the probability distributions for each of the data sets (bin size 0.001). Whilst the Cholesky data is generated from a uniform distribution, the resulting data set after factorisation resembles a normal distribution.

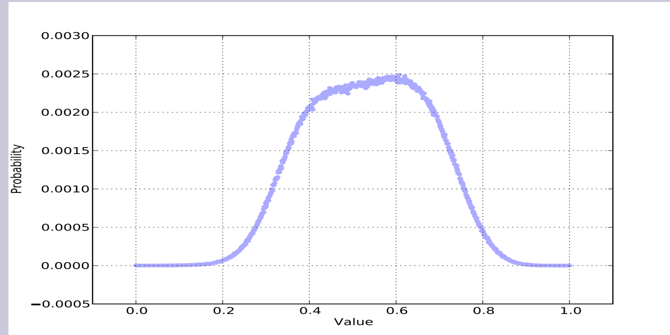
Uniform



Normal



Cholesky

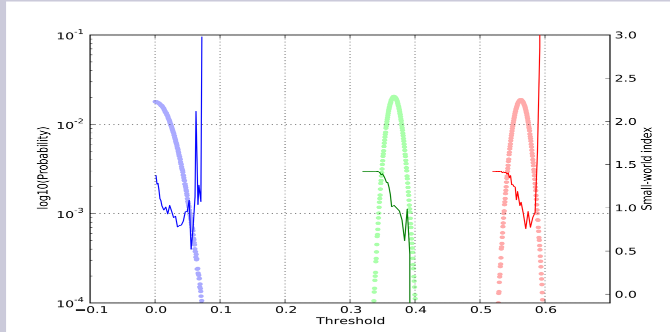


Small-world index

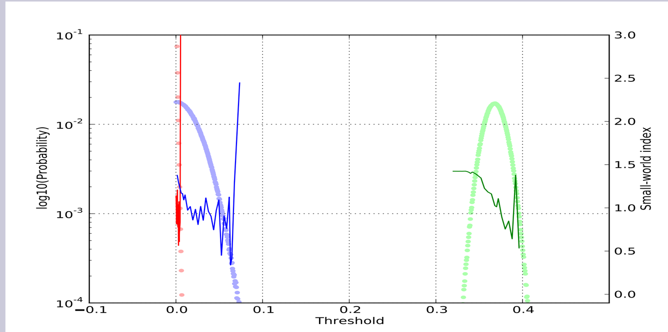
Shown below are the small-world indices for each of the data sets (the solid lines), and probability distributions (bin size 0.0005) for each of the three association measures (the scatter plots). Pearson correlation is plotted in blue, functional distance in green and coherence in red.

- Path lengths and clustering coefficients were approximated using the techniques described in [2] and [8], respectively.
- The distribution of values generated by each of the measures varies a lot for each of the data sets. This highlights the importance of choosing an appropriate measure, and threshold value, for graph generation.

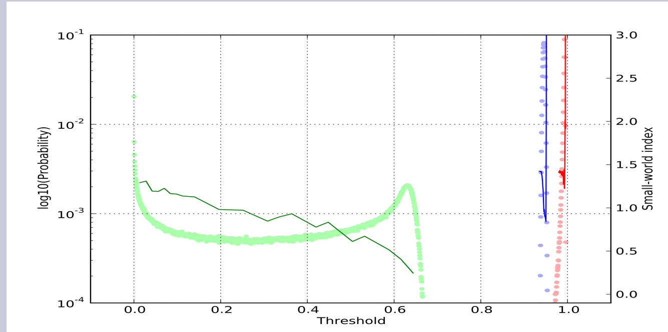
Uniform



Normal



Cholesky



Degree distribution

Shown below are degree distributions for graphs generated from each data set using each association measure. The red and blue scatter plots are from graphs generated with low and high threshold values respectively.

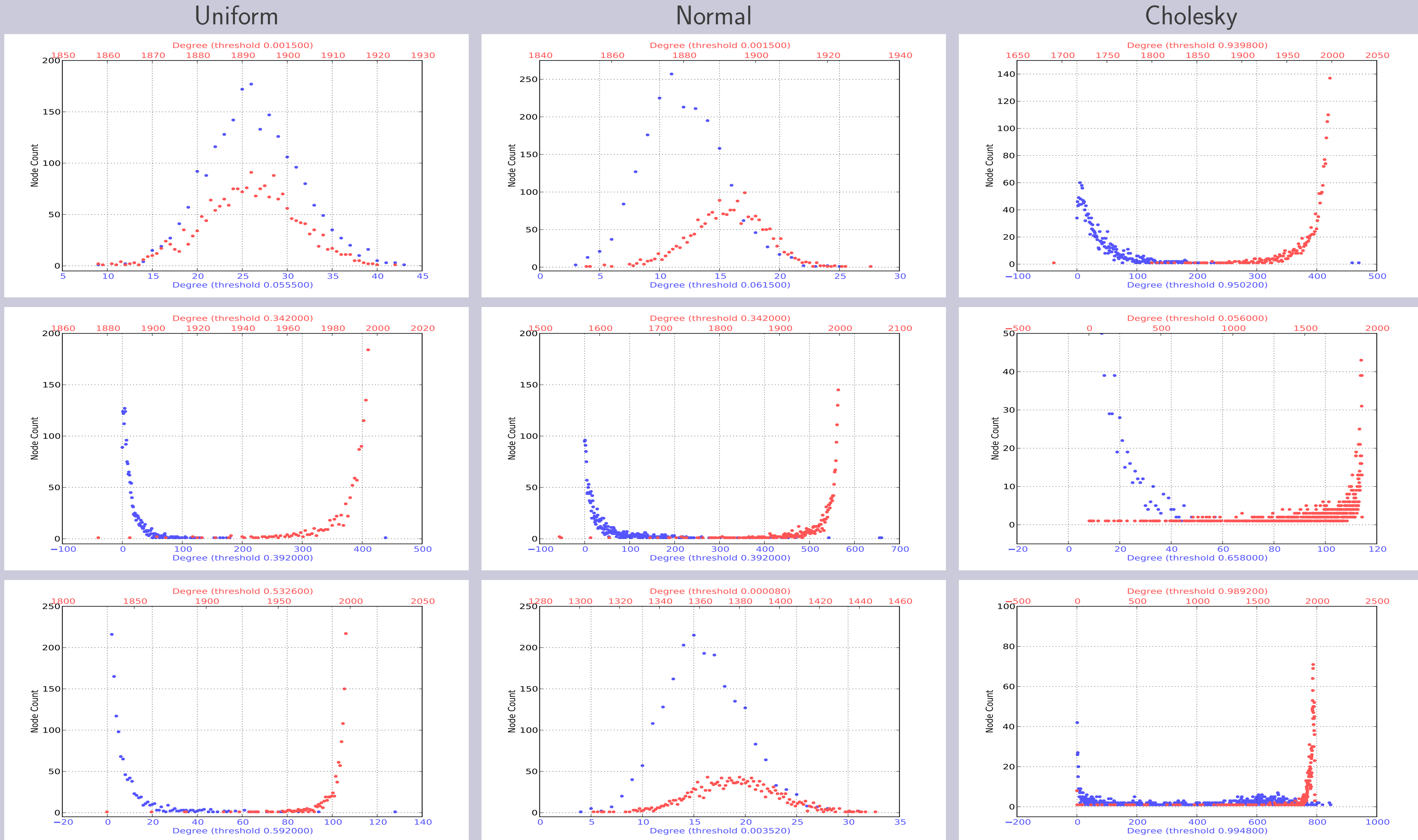
- Two types of degree distributions are produced; for example, using Pearson correlation upon the uniform data set produced a normal degree distribution. Contrast this with both functional distance and coherence, which produced exponential degree distributions on the same data set. An exponential degree distribution suggests a scale-free topology.
- All three association measures produced exponential distributions on the Cholesky generated data, highlighting the correlated nature of the data.
- Functional distance produced exponential distributions on all three data sets, and coherence produced an exponential distribution on the uniform data set. This suggests that these association measures may produce biased, or false-positive, results.

Uniform Normal Cholesky

Pearson Correlation

Functional Distance

Coherence



Pearson correlation coefficient

A normalised form of covariance; it is a standard measure of finding dependence between two variables. The original formula for calculating the coefficient divides the covariance by the product of the standard deviations of each variable to give a unitless value ρ between -1.0 and 1.0:

$$\rho(x, y) = \frac{\sum(x_i - \bar{x})\,(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2\,\sum(y_i - \bar{y})^2}}$$

A coefficient value near 0 indicates weak, or no association between the two variables, whereas a value near -1.0 or 1.0 indicates strong association. In this study, the absolute value of the correlation coefficient is used, yielding a range between 0.0 and 1.0.

Coherence

The coherence between two signals x and y is a real value between 0.0 and 1.0, which indicates association between the signals in the frequency domain. Coherence may be calculated over an entire signal, as in this study, or over discrete frequency bands. Coherence is calculated first by calculating the Fourier transform of each signal:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi k n i}{N}},\ k = 0, 1, \ldots, N-1$$

The power-spectral density is calculated by multiplying the Fourier transform of the signal by its complex conjugate:

$$P_{xx} = XX^* = |X|^2$$

The cross-spectral density of a pair of signals is calculated by multiplying the Fourier transform of one signal by the complex conjugate of the Fourier transform of the other signal:

$$P_{xy} = XY^*$$

The coherence may then be calculated as follows:

$$\frac{|\langle P_{xy}\rangle|^2}{\langle P_{xx}\rangle\langle P_{yy}\rangle}$$

As coherence measures association in the frequency domain, it is only suitable for use on data with relatively high temporal resolution. Use of coherence on e.g. fMRI data will not provide an accurate measure of association, as the poor temporal resolution of fMRI means that high frequency information is lost in the recording process. In contrast, coherence is ideal for use on EEG/MEG data, as these recording techinques are able to record at a high temporal frequency.

Functional Distance

The functional distance, proposed by [6], is the Euclidean distance between corresponding points from two time series. The functional distance $||x - y||$ between time series x and y is defined as:

$$||x - y|| = \sqrt{\sum(x_i - y_i)^2}$$

This is then embedded within a Gaussian kernel function:

$$w(x, y) = e^{-\left(\frac{||x-y||}{\sigma\xi}\right)^2}$$

where ξ is the median of functional distances between all pairs of nodes, and σ is a scaling factor ranging from 0.5 to 2.0. In this study, we have used $\sigma = 1.0$.

References

[1] Sophie Achard and Ed Bullmore. Efficiency and cost of economical brain functional networks. *PLoS Comput. Biol.*, 3(2):e17, 2007.

[2] Shinichi Honiden, Michael E. Houle, Christian Sommer, and Martin Wolff. Approximate Shortest Path Queries in Graphs Using Voronoi Duals. In *Proceedings of the 6th International Symposium on Voronoi Diagrams in Science and Engineering*, pages 53–62, 2009.

[3] Marie-Therese Horstmann, Stephan Bialonski, Nina Noennig, Heinke Mai, Jens Prusseit, Jörg Wellmer, Hermann Hinrichs, and Klaus Lehnertz. State dependent properties of epileptic brain networks: comparative graph-theoretical analyses of simultaneously recorded EEG and MEG. *Clin Neurophysiol*, 121(2):172–85, 2010.

[4] Viktor K. Jirsa and A. R. Mcintosh, editors. *Handbook of Brain Connectivity*. Springer Complexity. Springer, 2007.

[5] Maria Markosova, Liz Franz, and Lubica Benuskova. Topology of Brain Functional Networks: Towards the Role of Genes. *Advances in Neuro-Information Processing*, 5506:111–118, 2009.

[6] X Shen, X Papademetris, and R T Constable. Graph-theory based parcellation of functional subunits in the brain from resting-state fMRI data. *Neuroimage*, 50(3):1027–35, 2010.

[7] C.J. Stam and B. W. van Dijk. Synchronization likelihood: an unbiased measure of generalized synchronization in multivariate data sets. *Physica D*, 163(3-4):236–251, 2002.

[8] Dorothea Wagner Thomas Schank. Approximating Clustering Coefficient and Transivity. *Journal of Graph Algorithms and Applications*, 9(2):265–275, 2005.

[9] Martijn P van den Heuvel, Cornelis J Stam, René S Kahn, and Hilleke E Hulshoff Pol. Efficiency of functional brain networks and intellectual performance. *J. Neurosci.*, 29(23):7619–24, 2009.

[10] Liang Wang, Chaozhe Zhu, Yong He, Yufeng Zang, Qingjiu Cao, Han Zhang, Qiuhai Zhong, and Yufeng Wang. Altered small-world brain functional networks in children with attention-deficit/hyperactivity disorder. *Hum Brain Mapp*, 30(2):638–49, 2009.

http://miracle.otago.ac.nz/postgrads/pmccarthy/
http://www.cs.otago.ac.nz/homepages/lubica/
pauld.mccarthy@gmail.com
lubica@cs.otago.ac.nz