

# *Primary evolving networks and the comparative analysis of robust and fragile structures*

Article

Accepted Version

Grindrod, P., Stoyanov, Z. V., Smith, G. M. and Saddy, J. D.  
ORCID: <https://orcid.org/0000-0001-8501-6076> (2013) Primary evolving networks and the comparative analysis of robust and fragile structures. *Journal of Complex Networks*, 2 (1). pp. 60-73. ISSN 2051-1329 doi:  
<https://doi.org/10.1093/comnet/cnt015> Available at  
<https://centaur.reading.ac.uk/34017/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

Published version at: <http://comnet.oxfordjournals.org/content/early/2013/10/16/comnet.cnt015.abstract>

To link to this article DOI: <http://dx.doi.org/10.1093/comnet/cnt015>

Publisher: Oxford University Press

Publisher statement: This is a pre-copyedited, author-produced PDF of an article accepted for publication in *Journal of Complex Networks* following peer review. The definitive publisher-authenticated version is available online at:  
<http://comnet.oxfordjournals.org/content/early/2013/10/16/comnet.cnt015.abstract>

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

**CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

# Primary Evolving Networks and the Comparative Analysis of Robust and Fragile Structures

Peter Grindrod<sup>1</sup>, Zhivko V. Stoyanov<sup>\*1</sup>, Garry M. Smith<sup>2</sup>, and J. Douglas Saddy<sup>3</sup>

<sup>1</sup>Department of Mathematics and Statistics, Centre for the Mathematics of Human Behaviour (CMOHB), University of Reading, UK

<sup>2</sup>School of Systems Engineering, University of Reading, UK

<sup>3</sup>Centre of Integrative Neuroscience and Neurodynamics, University of Reading, UK

October 4, 2013

## Abstract

In this paper we consider the structure of dynamically evolving networks modelling information and activity moving across a large set of vertices. We adopt the *communicability* concept that generalizes that of centrality which is defined for static networks. We define the primary network structure within the whole as comprising of the most influential vertices (both as senders and receivers of dynamically sequenced activity). We present a methodology based on successive vertex knockouts, up to a very small fraction of the whole primary network, that can characterize the nature of the primary network as being either relatively robust and lattice-like (with redundancies built in) or relatively fragile and tree-like (with sensitivities and few redundancies). We apply these ideas to the analysis of evolving networks derived from fMRI scans of resting human brains. We show that the estimation of performance parameters via the structure tests of the corresponding primary networks is subject to less variability than that observed across a very large population of such scans. Hence the differences within the population are significant. **Keywords.** networks, communicability, brain science, fMRI data, robustness, bimodality.

## 1 Introduction

There is an increasing interest in evolving graphs: networks where edges appear and disappear over time [Grindrod and Higham, 2010; Crofts and Higham, 2011; Estrada, 2011]. Such networks

---

<sup>\*</sup>Corresponding author. *Email address:* z.v.stoyanov@reading.ac.uk

model a range of phenomena where information is communicated from vertex to vertex. The time ordering of the graphs (changing with the discrete appearance and disappearance of edges) induces an asymmetry; since if  $A$  communicates with  $B$ , and then later  $B$  communicates with  $C$ , the information from  $A$  can reach  $C$  but not vice versa. For this reason some generalizations of Katz centrality [Katz, 1953] have been developed so as to identify the role of individual vertices within evolving networks as influential sources of information or efficient sinks for information [Grindrod et al., 2011; Estrada et al., 2012; Grindrod and Higham, 2012]. These ideas and methods have been applied to large scale networks form a range of applications including social media, and peer to peer telecommunications and emails.

When evolving networks are very large (in terms of the number of vertices) there is usually some need to summarize those networks. Here we introduce the idea of dividing the network into a primary network containing all of the influential vertices and a consequent secondary network containing the less important vertices. We shall discuss how this differs from existing coarse grain-ing approaches [Mucha et al., 2010; Gfeller and Rios, 2007, 2008; Itzkovitz et al., 2005], which provide a different type of (mesoscopic) summary.

Having identified a large primary network we have a need to characterize and summarize its structure: is it relatively lattice-like and robust to insults, or is it relatively tree-like and fragile? We introduce a methodology to test for this by making successive knockouts of vertices within the primary network and examining the losses in overall functionality. By knocking-out very small fractions (typically 1% of the whole) we can remain within the linear regime and avoid second order, collaborative, loss-effects. Thus we can classify networks, from within a possibly large population of similar examples, with respect to performance measures<sup>1</sup> representing the relative size of the total functional loss due to a given number of knockouts, and the variability of those incremental losses.

To illustrate these ideas we consider their application to fMRI scans of (resting) human brains and compare the performance of the primary networks from almost 1000 such brains. We will show that the variation in the performance measures observed for an individual brain (due to analysing different random knockout sequences) is far less than the variation observed across a large population of brains. This leads to the inevitable conclusion that such individual brains are significantly different, and in particular that they can be relatively robust or relatively fragile when subject to successive knockouts of rather small scale components.

We present the framework to make this analysis tractable for primary networks of  $10^5$  or more vertices. The discussion of the application is relatively self-contained so as to make the paper accessible for analysts who can adapt and extend this methodology to any other types of data sets where *information* or *activity* of some kind is observed passing dynamically around a very large population of entities.

---

<sup>1</sup>Here, and elsewhere in the text, the phrase *performance measure* is used for descriptive purposes. Strictly speaking, the performance measures defined in this paper do not satisfy the usual measure and norm axioms.

## 2 Weighted evolving networks and communicability

An undirected weighted graph defined over a set of  $n$  vertices,  $V = \{v_i | i = 1, \dots, n\}$ , is such that for each possible edge, between  $v_i$  and  $v_j$  say, we have a real non-negative weight  $a_{ij}$ . The weighted adjacency matrix  $A$ , with its  $(i, j)$ -th term given by  $a_{ij}$ , is symmetric and non-negative, and is equivalent to a pairwise *similarity matrix* that might be used in clustering objects (here represented by the vertices). We shall always assume that the diagonal terms in  $A$  are all zero (so there are no self-connections).

Now consider an evolving weighted network given over  $K$  consecutive discrete time steps as a sequence of such undirected weighted graphs represented by a corresponding sequence of weighted adjacency matrices,  $\{A_1, A_2, \dots, A_K\}$ . A *dynamic path* from  $v_i$  to  $v_j$  is a sequence of successive adjoining edges, specified from the sequence of the adjacency matrices, linking  $v_i$  to vertex  $v_j$  through intermediate vertices, such that each edge occurs at the same time step or a later one than that of the previous edge. The time ordering of the sequence allows us to define the communicability matrix,  $Q$ , [Grindrod et al., 2011] as the ordered product of resolvents:

$$Q = (I - \eta A_1)^{-1} \cdot (I - \eta A_2)^{-1} \dots (I - \eta A_K)^{-1} = \prod_{k=1}^K (I - \eta A_k)^{-1},$$

where  $0 < \eta < 1/\max\{\rho(A_k)\}$  is a constant discount factor ensuring convergence (for more details on the role of, and possible choices for,  $\eta$ , see [Greetham et al., 2013] and the references therein). Each element of  $Q$  provides a sum over all possible vertex to vertex dynamic paths, of the products of the edge weights, discounted for length. This is a generalization of Katz centrality (which applies for a static network, effectively recovered here when  $K = 1$ ).

It is evident that  $Q$  is generally not symmetric, because of the time ordering of the sequence. The  $i$ -th row sum of  $Q$  represents all of the paths emanating from  $v_i$ , and is a measure of  $v_i$ 's power as a *source* or initiator of dynamic pathways; and  $i$ -th column sum of  $Q$  represents all of the paths coming into  $v_i$ , and is a measure of the  $v_i$ 's power as a *sink* or destination for dynamic pathways. Intuitively if many dynamic paths go through some vertex,  $v_i$ , then the upstream and downstream contributions of those paths are counted within the corresponding column and row sums of  $Q$ . Hence those vertices lying on the *main highways* of dynamic propagation (highly weighted paths) will have relatively large corresponding row and column sums.

We will write the  $n$ -vector of the row sums as  $\mathbf{b}$ , and the  $n$ -vector of the column sums as  $\mathbf{r}$ , given by

$$\mathbf{b} = (b_1, \dots, b_n)^T = Q\mathbf{1}, \quad \mathbf{r} = (r_1, \dots, r_n)^T = Q^T\mathbf{1}.$$

The particular case we have in mind here is where  $n$  is rather large. In that case we might seek to avoid calculating  $Q$  directly, but instead we are able to directly calculate  $\mathbf{b}$  and  $\mathbf{r}$  by multiplying  $Q$  by  $\mathbf{1} = (1, 1, \dots, 1)^T$  from the left or the right, respectively. In practice we start out from the estimates  $\tilde{\mathbf{r}} = \mathbf{1}$  and  $\tilde{\mathbf{b}} = \mathbf{1}$ , and update those via successive linear solves equivalent to multiplication by the respectively, forward and backward ordered, resolvents. We might also wish to avoid calculating or holding the  $A_k$ s in memory. For example, if we can write  $A_k = X_k \cdot X_k^T$  where  $X_k$  is  $n \times m$  with  $m \ll n$ , then we may work with the  $X_k$ s rather than the  $A_k$ s.

## 2.1 Nested primary networks

Consider the distribution of the values contained in  $\mathbf{b} = (b_1, \dots, b_n)^T$ , and respectively  $\mathbf{r} = (r_1, \dots, r_n)^T$ . This may be bimodal or multimodal within some applications, and thus there may be natural *break-points* within these distributions. In any case let us suppose that for each distribution we may set a suitable threshold value  $\beta^*$ , and respectively  $\rho^*$ , above which we consider vertex specific  $b$ -values an  $r$ -values to be significant; and below which we consider them to be negligible.

For any given pair of non-negative values  $(\beta^*, \rho^*)$ , we shall define the associated *primary network* as that consisting of all edges at all time steps connecting only pairs of vertices within the set  $V^*$ :

$$V^*(\beta^*, \rho^*) = \{v_i \mid b_i \geq \beta^* \text{ or } r_i \geq \rho^*\} \subset V.$$

As either  $\beta^*$  or  $\rho^*$  increases  $V^*$  becomes smaller, approaching  $\emptyset$  once both  $\beta^* > \max b_i$  and  $\rho^* > \max r_i$ . Thus we may generate nested primary networks defined as both  $\beta^*$  or  $\rho^*$  increase.

In practice in considering the primary network, the subgraph induced by the evolving network in restricting  $V$  to  $V^*$ , we are excluding those edges which, even if they have a relatively large weight in some or other time steps, merely connect vertices that are not significant as either sources or sinks of the dynamic paths within the full evolving network system.

For the reduced, primary network, defined on the vertices in  $V^*$ , we can calculate the associated communicability,  $Q^*$ , and thus the associated  $\mathbf{b}^*$  and  $\mathbf{r}^*$ . These are the descriptors of that primary network on  $V^*$ . Only those edges connecting those vertices in  $V^*$  are admissible. This is done simplest by removing all rows and columns in the  $A_k$ 's whenever the corresponding vertex  $v_i$  is in  $V \setminus V^*$ .

Intuitively, for primary networks that are meaningful, we expect the resulting values within  $\mathbf{b}^*$  and  $\mathbf{r}^*$  to be highly correlated with their precursors for the full network,  $\mathbf{b}$  and  $\mathbf{r}$ . This is so because the counts of pathways made within  $\mathbf{b}$  and  $\mathbf{r}$  will be dominated by paths within the primary network. Clearly, the existence of a meaningful primary subnetwork depends on the structure of the whole network, and also on the values of  $\beta^*$  and  $\rho^*$ . There certainly are instances of networks that do not contain meaningful primary subnetworks. However, we do not concern ourselves with this issue here. Instead, the main purpose of this paper is to highlight a technique for extracting subnetworks, which are considerably smaller in size than the original network, but whose contribution to the overall communication is significant.

The identification of primary (sub)networks serves a very distinct purpose from the concept of coarse graining [Mucha et al., 2010]. Here the aim is to identify the main subnetwork(s) that gives rise to the majority of pathways, as counted by the communicability (a dynamic form of centrality). We will disregard individual edges for which at least one of the end points (nodes) has both a relatively low source and sink communicability score. This would mean that such edges lead nowhere and merely play a role in relatively few pathways, even if they possess a high weighting. In contrast methods of coarse graining seek a macroscopic representation of the whole network by replacing subcomponents of the network's vertices with single "meta" vertices; and then inducing some aggregated weightings for connections between those meta subcomponents. Both are simplifications of complex networks, but they achieve rather different things.

## 2.2 Associated secondary networks

Consider a given primary network containing all edges, at each of the time steps, connecting pairs of vertices within  $V^*(\beta^*, \rho^*)$ . The associated secondary network consists of all of the edges, at each time step, that are not included within the primary network. Such edges must connect at least one vertex which has both sub-threshold row and column sums within the full communicability matrix, and thus is not in  $V^*(\beta^*, \rho^*)$ .

## 2.3 Probing the structure of primary networks

Given a (large) evolving primary network with more than 100000 vertices, we wish to investigate the nature of its structure through sampling rather than exhaustive analysis. We propose to do so by making a number of sequential *knockouts*. In practice this means removing vertices, one at a time, from the network and then recalculating a quantity that measures the primary network's functionality at every successive iteration.

We proceed as follows. Communicability matrices, such as  $Q$  and  $Q^*$ , summarise the functionality observed in an evolving network and are nonnegative. We define the norm of such matrices as the sum of their elements

$$\|Q\| = \mathbf{1}^T Q \mathbf{1} = \mathbf{1}^T \mathbf{b} = \mathbf{1}^T \mathbf{r}.$$

Now consider a generated sequence of communicability matrices  $\{Q_j | j = 0, \dots, M\}$ , where  $Q_0 = Q^*$ , that for the original primary network, and then successive elements are generated by randomly selecting a vertex that is "live" within the previous network, and deleting it. At each iteration we may recalculate the communicability. Here  $\eta$  remains fixed. Thus as vertices, and hence edges, and consequently some paths, are deleted, we obtain a sequence  $\{Q_j\}$  such that the corresponding sequence of norms  $\{\|Q_j\|\}$  is monotonically decreasing. We shall only knockout  $M$  vertices, which is ideally less than 1% of all vertices in  $V^*$ ; so that there is a low probability of deleting vertices that are highly connected together. We desire that the degradation remains firmly within the linear regime (with small numbers of independent knockouts). We shall observe the step-by-step degradation of the evolving network, as measured by the monotonic reductions in  $\{\|Q_j\|\}$ . For some values of  $j$  we will annihilate a vertex that plays little role in many paths with the result that  $\|Q_{j-1}\| - \|Q_j\|$  is relatively small. For other values we may annihilate a vertex with a large communicability score and hence  $\|Q_{j-1}\| - \|Q_j\|$  will be relatively large.

Suppose the evolving primary network is very lattice-like, with a high Watts-Strogatz clustering coefficient, say. Then there is a large amount of redundancy in such a primary network, and since the lattice is relatively homogeneous almost every random knockout will produce a similar reduction in functionality, and the overall progress will be close to linear. For example, imagine a network on a grid, like the roads of Manhattan. If we knock out almost any intersection, the traffic can drive two further blocks around it and little functionally is lost. On the other hand, suppose the primary networks is very tree-like, with few cycles of any length. Then, when some vertices with high centrality are knocked-out, we would expect to see a large reduction in functionality. Think of the UK railway network, for example. If we knockout Birmingham New Street the network loses a large amount of functionality, yet if we knockout Henley-on-Thames virtually nobody will notice.

The random degradation process is also suggestive of early stage decline or damage of an ageing network. This analogy is particularly useful in considering human brains of course, where early onset cognitive decline is a major issue of interest. In fact it is clear that we ought to see a range of different experiences of cognitive degradation displayed within ageing populations. Some older people lose cognitive functionality in occasional, but large, steps (presumably having occasional critical, un-replaceable, catastrophic losses), hence being somewhat *fragile*. Yet some people's loss is long term, and relatively smooth and slow (presumably exploiting some network redundancies, and hence displaying a functional robustness to ageing). Thus the proposed approach is a useful way to (destructively) test the network, as well as providing an experimental analogue to random degradation through ageing. It may form a basis for a future clinical analysis of fMRI scans.

The nature of the degradation arising from a random sequence of knockouts of these networks may be characterized by two performance measures: the size (in absolute terms) and the nature (variability) of the sequential reductions in functionality, as measured by the  $\{\|Q_j\|\}$ .

Let  $M$  be the number of vertices removed. Then we calculate the quantity

$$\left(1 - \frac{\|Q_M\|}{\|Q_0\|}\right),$$

which is the fractional loss of functionality (as a result of  $M$  voxel knockouts). If this is small then  $M$  insults have had little impact on the primary network, which must consequently be relatively large. If this is large then the  $M$  insults have removed a more significant amount of functionality and the primary network must consequently be relatively small.

Next consider the successive fractional losses,

$$\left\{ \frac{\|Q_{j-1}\| - \|Q_j\|}{\|Q_0\| - \|Q_M\|} \mid j = 1, \dots, M \right\},$$

and suppose that they are sorted into descending order. Then we may plot the cumulative fraction of total loss against the cumulative fraction of the total knockouts (ordered by descending size of loss), see Fig. 1. This curve lies within the unit square, connecting  $(0,0)$  to  $(1,1)$ , above the diagonal with a negative second derivative. We shall calculate the area under this (ROC-like) curve. It is equal to one half if and only if all of the fractional losses are equal (all knockouts produce the same loss). It is equal to exactly one if and only if all of the fractional losses are zero, except for one which is unity (a single knockout accounts for all of the loss). Heuristically, we may say that if this area is small, and close to a half, then the evolving primary network appears to be lattice-like, with many redundancies, and is thus robust; if the area is larger, then the evolving primary network has less redundancy and robustness and is more tree-like, and hence is relatively fragile. We shall consider this pair of performance measures, plotted as a point in the plane, as a summary of the primary network's structure.

Now if we do this calculation many times we will obtain distinct results due to the random selection of successive knockouts. Thus, by resampling the knockout sequence many times over, we calculate an estimate for the means for both measures, together with estimates for the corresponding ranges sampled on either side, see Fig. 2. Hence, we might compare a collection of distinct



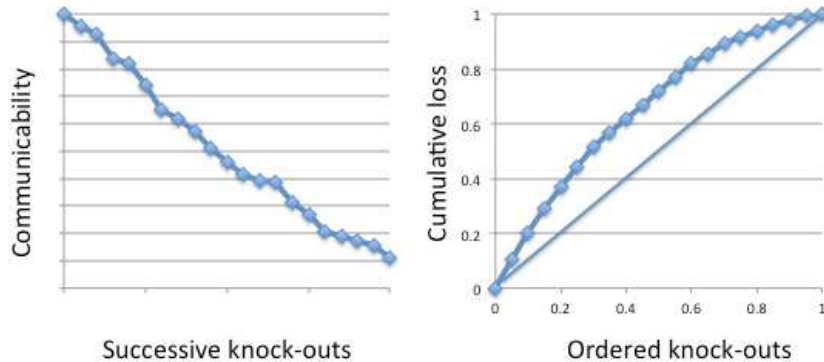


Figure 1: The effect of sequential vertex knockouts from the primary networks:  $\|Q_j\|/\|Q_0\|$  versus knockout  $j$  (left); and the cumulative distribution of loss versus the cumulative distribution of knockouts, sorted in descending order of size (right).

evolving primary networks via two performance measures: their (estimated) expected point location, together with their corresponding ranges (achieved with a given number of samples). It will be clear when the variability across the collection of individual primary networks is significantly larger than the sampling error ranges on the point estimates.

### 3 Primary networks for fMRI brain scan data

We consider data from an fMRI scan of a human brain, which contains around  $n = 2.5 \times 10^5$  voxels (small three dimensional volumes within which activity can be measured), which we shall treat as vertices. Here  $a_{ij}$  represents a one-sided covariance of the measured activities (transient blood oxygen level which is related to energy usage) within voxels  $v_i$  and  $v_j$ , over 10 successive time frames (from the scan). We step the 10-frame window through a full set of 110 time frames, producing an evolving weighted network over  $K = 11$  discrete time steps, as a sequence of such undirected weighted adjacency matrices.

The calculation of the  $n$  by  $n$  communicability matrix  $Q$  represents an immediate challenge. We get around this by noting that each weighted adjacency matrix,  $A_k$ , can be represented by the (outer) product  $X_k X_k^T$ , where  $j$ -th row of the matrix  $X_k$  contains the activities of voxel  $j$  over all snapshots contained in time step  $k$ ; plus a small correction that takes care of the diagonal elements. Therefore, we do not need to store any massive adjacency matrix and we may use Taylor expansion to compute  $(I - \eta A_k)^{-1} \mathbf{x}$ . In this way we control the precision of the approximation and the computational cost. Finally, in order to estimate a suitable value for  $\eta$ , we may compute the largest eigenvalue of  $A_k$  via the Power method (c.f. [Golub and Loan, 1996]), again without holding those matrices. Typically we set  $0 < \eta < 0.25 / \max\{\rho(A_k)\}$  in order to ensure the convergence of the resolvents.

Now we can visualize the roles played by the distinct voxels within dynamic pathways (those including edges from two or more timesteps). For example, if we take the *source-sink* difference,  $\mathbf{b} - \mathbf{r}$ , we can eliminate the counts of all vertex-to-vertex paths that take place within any single timestep (since such paths are reversible, and opposite, so they contribute to both counts). Then

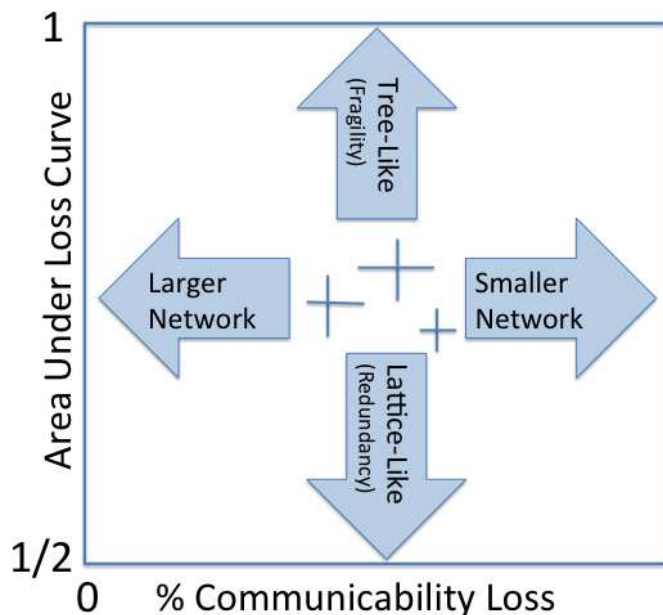


Figure 2: Three primary networks plotted as estimates and two sided ranges, with respect to both performance measures.

we may see those regions of the brain, voxel by voxel, for which  $\mathbf{b}$  dominates  $\mathbf{r}$ : that is, they have *more* downstream paths than upstream paths, coloured red in Fig. 3. Similarly, those voxels for which  $\mathbf{r}$  dominates  $\mathbf{b}$ : that is, they have *more* upstream paths than downstream paths, which are coloured green in Fig. 3.

Clearly these dynamical paths (representing successive chains of events carrying over at least two time-steps) yield a highly structured field. Moreover, if we randomly permute all of the time-steps (permute the  $A_k$ 's) and then repeat the whole operation, the resulting differences,  $\mathbf{b} - \mathbf{r}$ , become much smaller. Such a permutation can be carried through to show this field observed within the unpermuted data is highly statistically significant. So the dynamical information extracted confidently reflects some sorts of processes that are actually taking place and is not simply an artifact of the observations or the method. The structures in Fig. 3 themselves are interesting too. They have relatively short wavelength and display clear striping throughout the cortex.

Scientists working in the fMRI brain scan field may have never encountered striping like this either because they are in the habit of defining static networks, where the communicability (centrality) matrix is symmetric and hence  $\mathbf{b} = \mathbf{r}$ , or else of analyzing the data at lower resolutions. A common reaction is to declare that this is *merely noise*, presumably because it shows evidence of dynamic structure within regions that they typically wish to “parcellate”, and is an *inconvenient* phenomenon. In fact these patterns are very far from being spatial noise indeed, and they have a very distinctive scale. Our permutation tests also show that the patterning is not the result of temporal noise: these patterns represent dynamical flows from small scale volumes behaving as relative sources and relative sinks for inter-brain communication.

The resulting distributions for  $\mathbf{b} = (b_1, \dots, b_n)^T$  and  $\mathbf{r} = (r_1, \dots, r_n)^T$  are shown in Fig. 4. From

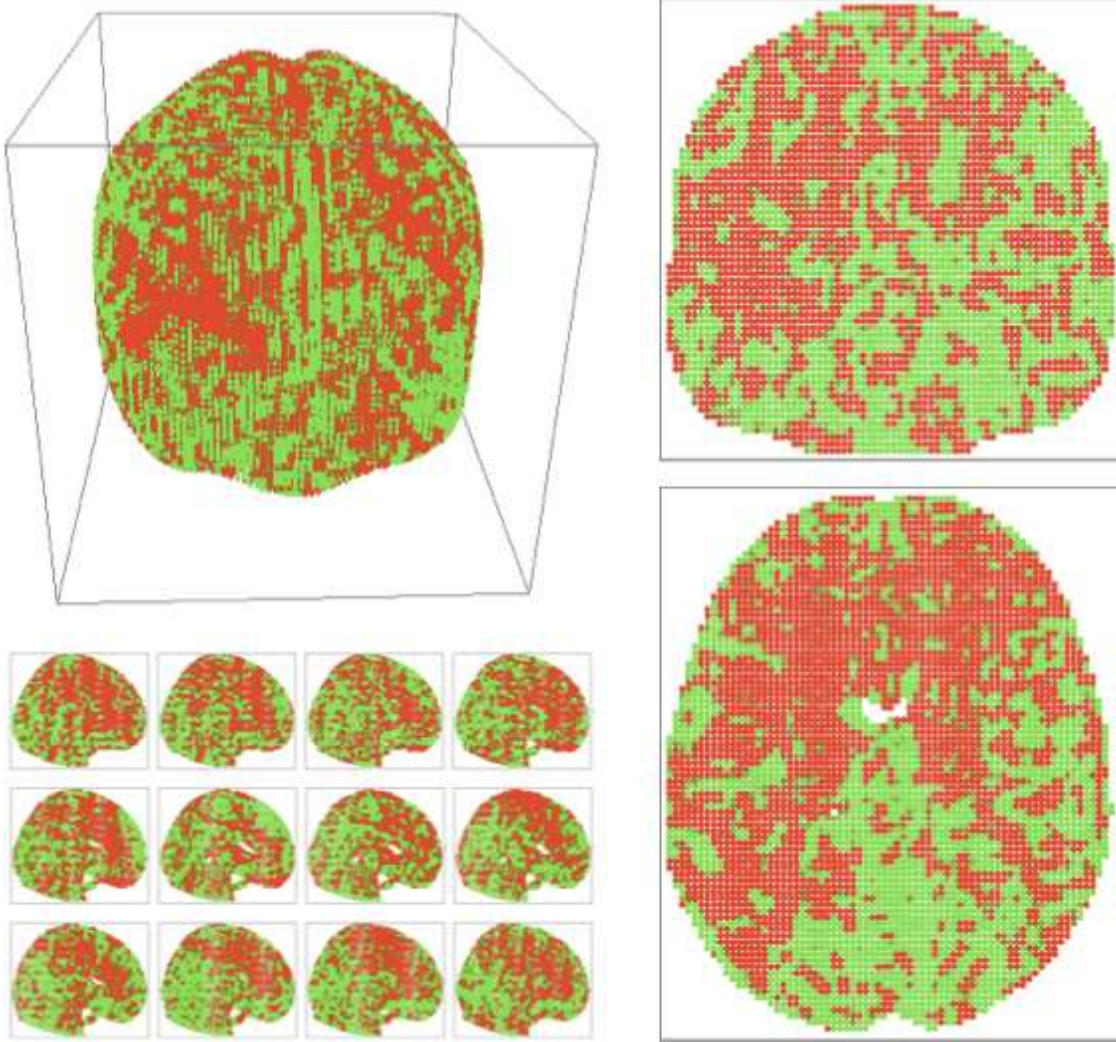


Figure 3: A 3D map of a brain obtained from the source minus sink scores,  $\mathbf{b} - \mathbf{r}$ .

these we select threshold values of  $\beta^*$  and  $\rho^*$  so as to retain the upper modes within the primary network. This means that approximately half the vertices ( $1.25 \times 10^5$ ) are retained within  $V^*$ .

Using this approach, we have analysed 967 separate fMRI scans, which are part of the data available from the 1000 Connectome Project<sup>2</sup>. The multimodal structure in these distributions is similar in all cases: so it is straightforward to select a primary network containing about half of the voxels.

Next we recalculate the measures associated with the primary network's communicability matrix,  $Q^*$ . In Fig. 5 we show the values obtained in  $\mathbf{b}^*$  versus those in  $\mathbf{b}$ ; and the values obtained in  $\mathbf{r}^*$  versus those in  $\mathbf{r}$ . Since the primary network is dominant within the full communicability matrix, by construction, these are very closely correlated.

To visualize the resulting primary network on the reduced set of vertices,  $V^*$ , consider the field

<sup>2</sup>For more information visit [http://fcon\\_1000.projects.nitrc.org/](http://fcon_1000.projects.nitrc.org/) or <http://www.nitrc.org/>.

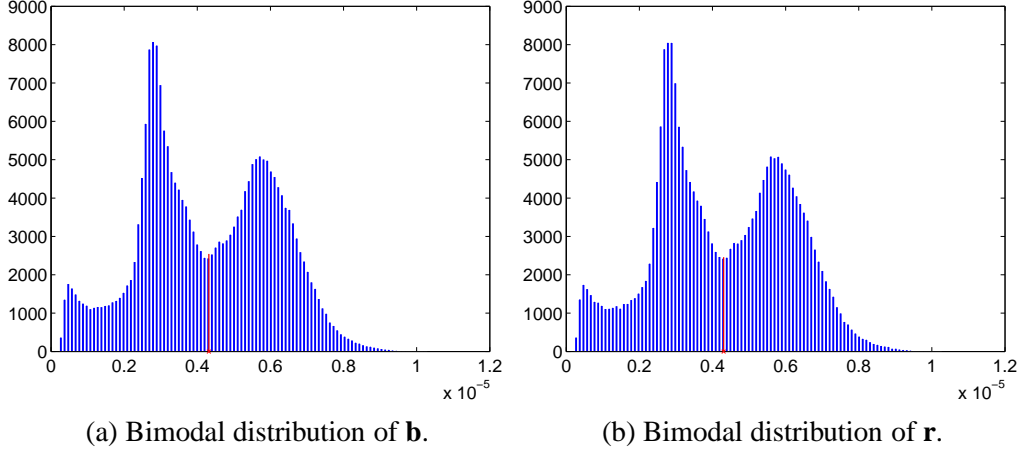


Figure 4: Histograms of the bimodal distributions of  $\mathbf{b}$  and  $\mathbf{r}$ . We have also computed and plotted the corresponding thresholds (red dashed line).

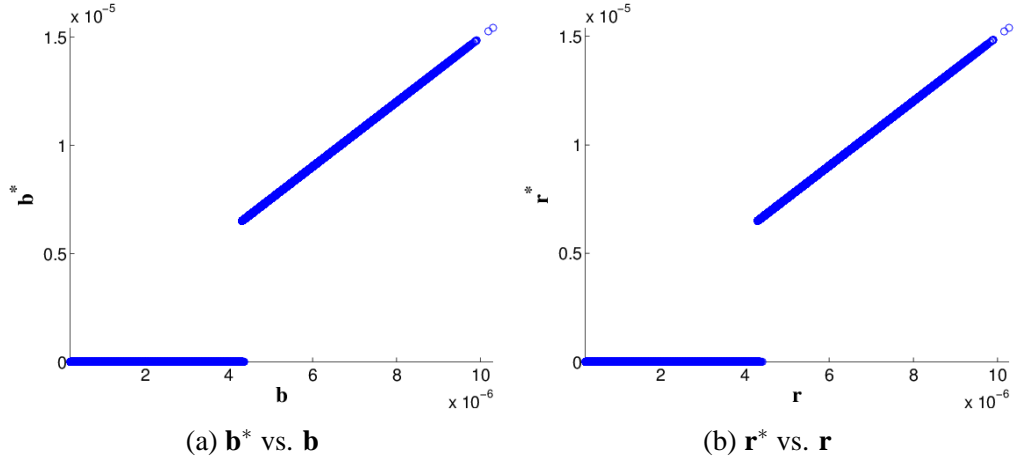


Figure 5: Scatter plot of  $\mathbf{b}^*$  vs.  $\mathbf{b}$  and  $\mathbf{r}^*$  vs.  $\mathbf{r}$ .

given by the source communicability, the row sums,  $\mathbf{b}^*$ . This is shown in Fig. 6. Notice that the left and right hemispheres have now become mostly separated within the primary network and there are some voids within the brain mass. The most extreme positive  $\mathbf{b}^*$ -values are towards the outside layers of the cortex.

Next we apply the method given in section 2.3 to consider an ensemble of 967 fMRI brains scans. These are all scans of resting brains, from a number of laboratories, and each has been downloaded from the connectome database and then normalized (mapped onto a standard voxelated representation). We also restricted each normalized scan to 110 time frames, and thus  $K = 11$  timesteps.

For each brain we proceeded independently as follows:

- (a) we identified the primary network using suitable threshold parameters  $(\beta^*, \rho^*)$ ;

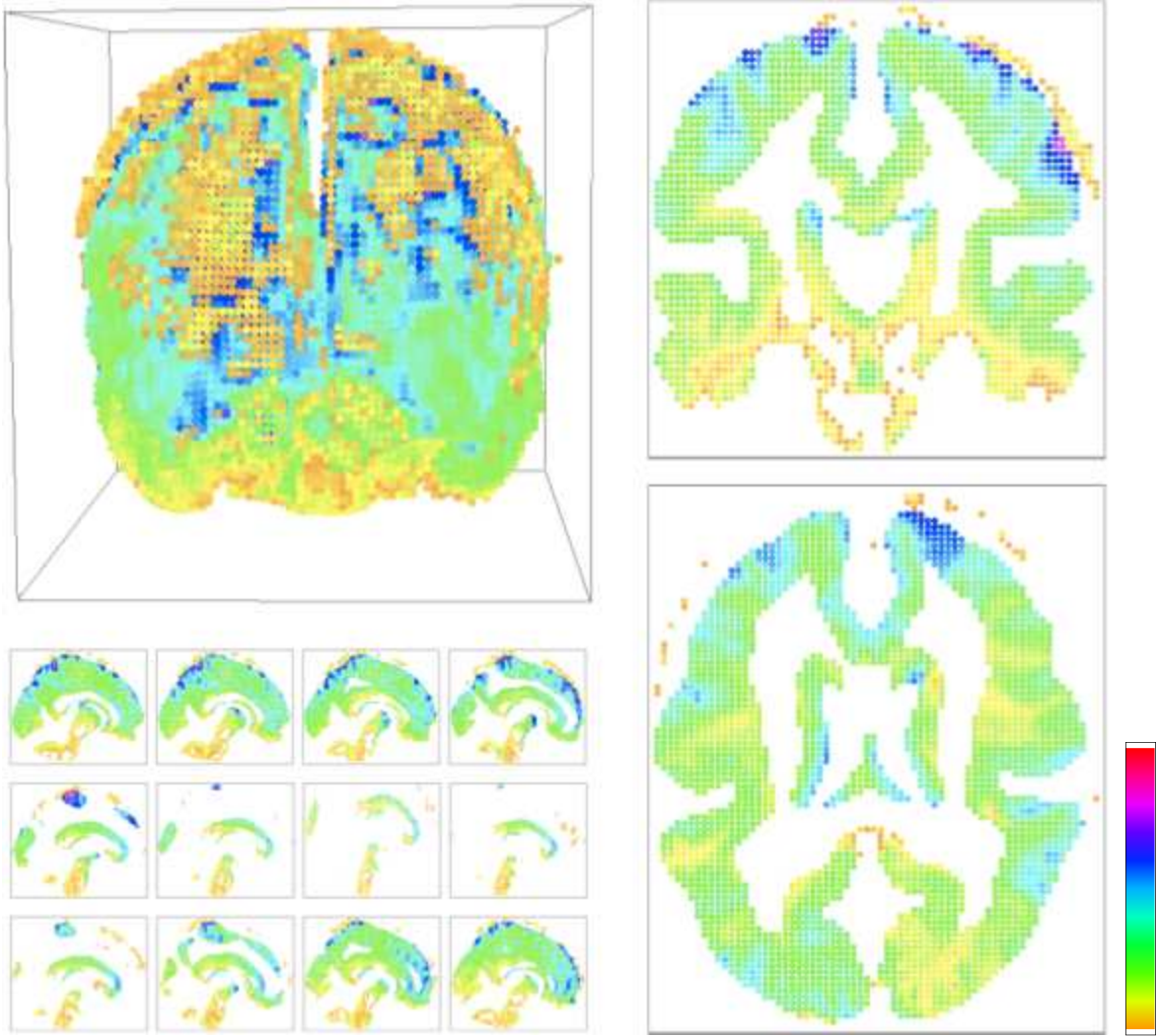


Figure 6: A 3D map of the  $\mathbf{b}^*$  measures for the primary network within the brain.

- (b) we calculated  $\|Q^*\|$  and its related measures (and tests);
- (c) we degraded the primary network with  $M = 1000$  successive voxel knockouts;
- (d) we repeated step (c) independently 100 times to estimate means and ranges for the two performance measures.

This process involved making around 100 000 separate communicability calculations for original and degraded primary networks; each of which, conceptually at least, was made based on over 11 (evolving) weighted graphs containing more than 100 000 vertices. Such communicability calculations were all made using the code given in [Stoyanov et al., 2013], executing across a hybrid cloud (a local physical compute cluster, as well as in Virtual Machines running in the Amazon AWS Infrastructure-as-a-Service cloud).

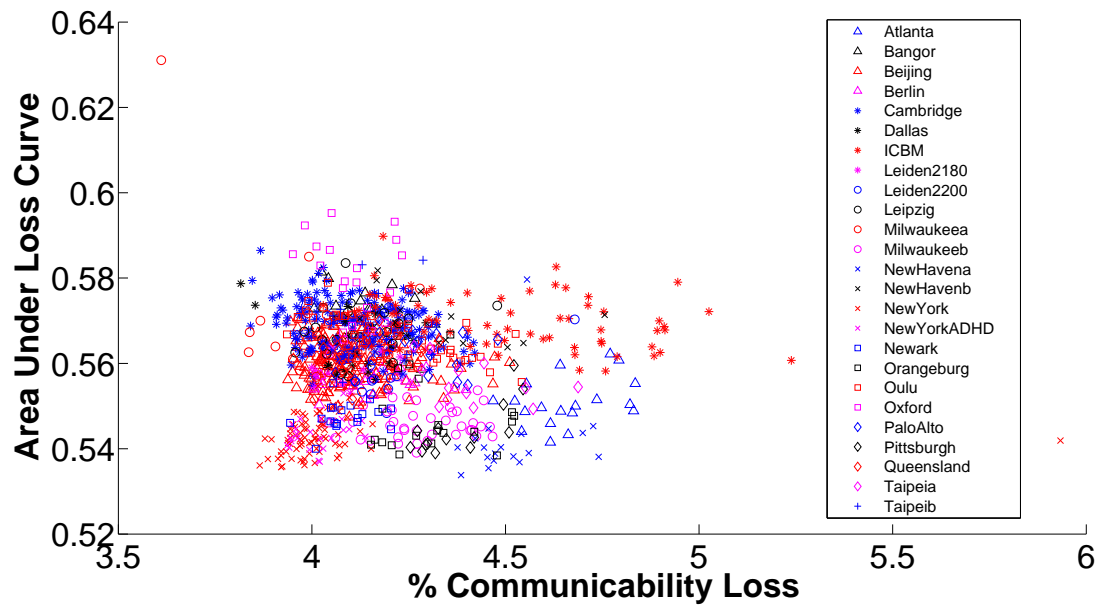


Figure 7: Scatter plot showing performance measures under degradation for 967 brains, indexed by investigating laboratory

In Fig. 7 we show a scatter plot of the results for all 967 brains from a number of investigating laboratories (see key). The variation across that population is far greater than the variations observed for individual brains when the degradations are resampled. To see this, consider Fig. 8, where for each of the brains from five of the laboratories we show the full range of variation in the performance measures achieved over 100 independent randomly sampled degradations. In all cases differences between some of the individuals' brains is greater than the corresponding individual degradation sampling errors.

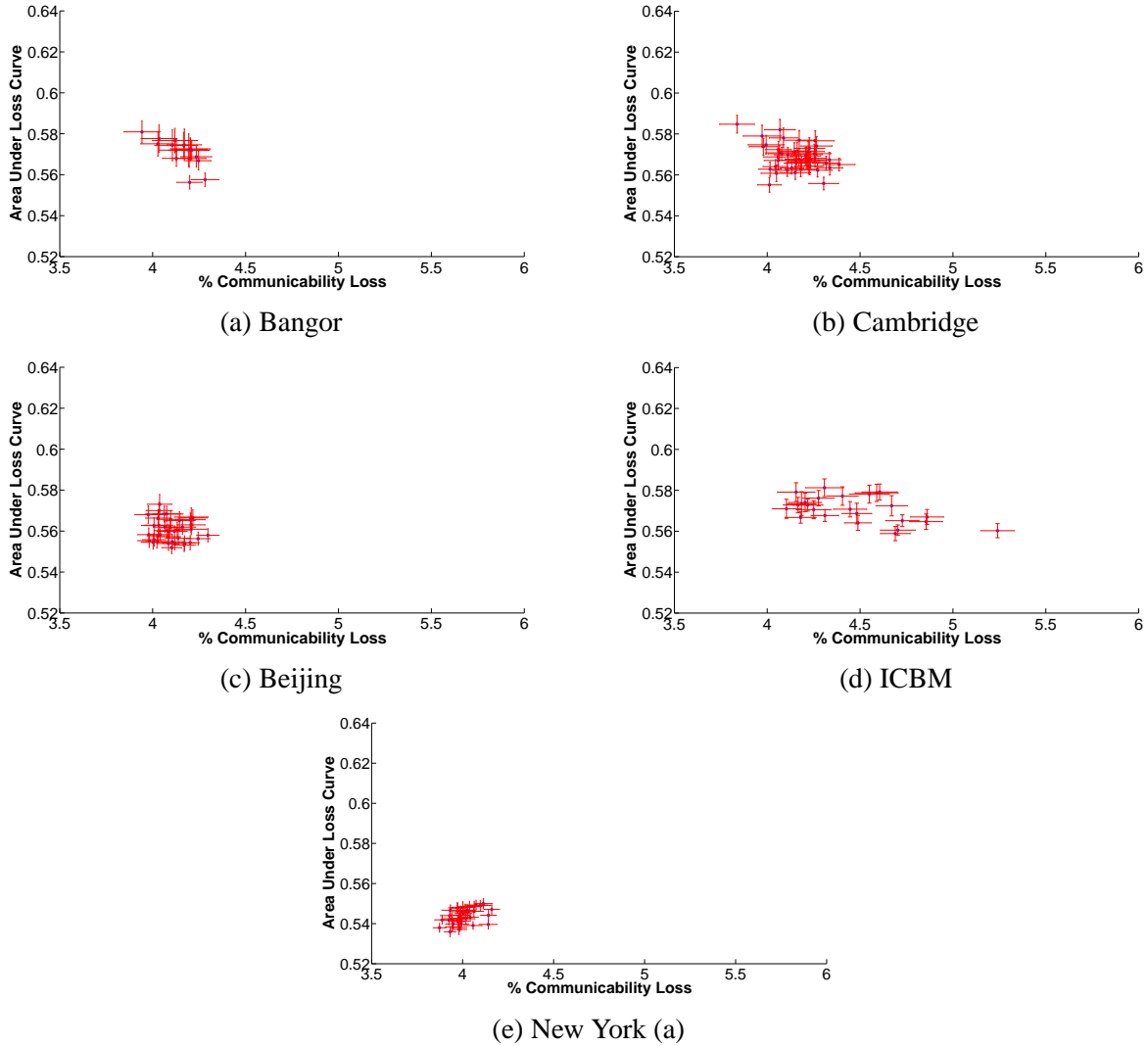


Figure 8: Individual brain performance measures with extreme ranges

## 4 Conclusions

For any vast evolving network the primary network represents those vertices (and their connecting evolving edges) which dominate the distribution of all possible dynamic pathways between all pairs of vertices. Coarse graining approaches [Mucha et al., 2010; Tozzini, 2005; Gfeller and Rios, 2007, 2008; Itzkovitz et al., 2005] summarize vast (static or dynamic) networks by introducing an intermediate, mesoscopic level representing components of the whole network, as single mesoscopic vertices connected appropriately so as to represent the microscope edges between vertices within each component. In many applications this is entirely appropriate. Here we have introduced the idea that if we wish to stay at the high resolution (microscopic scale), we might reduce the size of the network by retaining only those vertices and edges which might carry the major components

of any flow of information, or coherent behaviour. This is related to generalized Katz centrality (extended to evolving networks) rather than time dependent degree or frequency of edges. In some applications there may be a natural divide between primary and secondary networks, and we have illustrated this construction with one such application. The importance of this is that too often, very high resolution data, such as fMRI blood oxygen level images, available from modern powerful scanners, are reduced to analysis interaction between 500 or so parcels of vertices (defined in some way) so as to make analysis tractable. The fine resolution is thus lost. In fact the dynamical element of the scans is also often lost, with single static networks being extracted from the time dependent behaviour. Thus the concept of some vertices acting as sources and some as sinks for communication is simply unavailable in many of the published analyses.

The probing of the structure of primary networks via simulated (sampled) degradation, opens up a number of possibilities for future work and exploitation. In any application it is essential to show that the variation of performance measures due to sampling of knockouts is less than that observed across large populations of similar networks, as we have shown in the application here. Here we have suggested just two conceptually independent ways of measuring the performance and structure of primary networks (through simulated degradation): one measuring the size of impact and the other measuring the fragility/robustness of the network to insults. There may be other, more illuminating measures to be defined, and this issue is far from complete. Nevertheless, it cannot be argued that what we have seen is an artifact of the data or the analytics: it is lost when one permutes the time-steps (breaking the dynamics); the source-sink structure observed is far from random (and is in fact observable within all brains; yet individuals are distinct - like fingerprints).

The emergence of high resolution data in many fields demands that the analytics respect that, and we should not necessarily aggregate the activity into larger sets of mesoscopic voxels *a-priori*. To do so would lose and distort the structures presented here. Once we upscale (coarse grain) they are lost. Similarly the time resolution of vast data sets from many fields will become finer in the future and the analysis presented here could scale with such developments. It is fortunate indeed that this data deluge coincides with the availability of cloud-based, parallel, low cost, computing facilities for all.

## Acknowledgements

This research would not have been possible without the support of members of the Centre for Integrative Neuroscience and Neurodynamics (CINN) as well as the Centre for the Mathematics of Human Behaviour, both of which are at the University of Reading; Tamsin Lee who patiently prepared the data sets; Danica Greetham and Jon Ward; Tom Johnstone and Bhisma Chakrabarti who encouraged the application of communicability ideas to fMRI data; and our long-term collaborator Des Higham, who with Peter Grindrod and others developed the original ideas of communicability as a generalization of centrality for evolving graphs, and who has provided challenge and advice on many occasions.

This work benefited from funding, including the provision of direct support for Zhivko Stoyanov, from a number of EPSRC grants over the last three years: EP/F033036/1, Cognitive Systems



Science (Bridging the Gaps); EP/I016856/1, NeuroCloud; EP/H024883/1, An integrated neural field computational model; and EP/G065802/1, The Digital Economy HORIZON Hub.

## References

- J. J. Crofts and D. J. Higham. Googling the brain: Discovering hierarchical and asymmetric network structures, with applications in neuroscience. *Internet Mathematics (Special Issue on Biological Networks)*, 7(4):233–254, 2011.
- E. Estrada. *The Structure of Complex Networks: Theory and Applications*. Oxford University Press, 2011. ISBN 9780199591756.
- E. Estrada, N. Hatano, and M. Benzi. The physics of communicability in complex networks. *Physics Reports*, (514):89–119, 2012.
- D. Gfeller and P. De Los Rios. Spectral coarse graining of complex networks. *Phys. Rev. Lett.*, 99(3), 2007. doi: 10.1103/PhysRevLett.99.038701. URL <http://dx.doi.org/10.1103/PhysRevLett.99.038701>.
- D. Gfeller and P. De Los Rios. Spectral coarse graining and synchronization in oscillator networks. *Phys. Rev. Lett.*, 100(17):174104, May 2008. doi: 10.1103/PhysRevLett.100.174104.
- G. H. Golub and C. F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, third edition, 1996.
- D. V. Greetham, Z. V. Stoyanov, and P. Grindrod. On the radius of centrality in evolving communication networks. Submitted to the Journal of Combinatorial Optimization (JOCO), 2013.
- P. Grindrod and D. J. Higham. Evolving graphs: Dynamical models, inverse problems and propagation. *Proc. R. Soc. A*, 466(2115):753–770, 2010.
- P. Grindrod and D. J. Higham. A matrix iteration for dynamic network summaries. *SIAM Review*, 2012. ISSN 1095-7200. In press.
- P. Grindrod, D. J. Higham, M. C. Parsons, and E. Estrada. Communicability accross evolving networks. *Phys. Rev. E*, 83(4), 2011. ISSN 1539-3755. doi: 10.1103/PhysRevE.83.046120.
- S. Itzkovitz, R. Levitt, N. Kashtan, R. Milo, M. Itzkovitz, and U. Alon. Coarse-graining and self-dissimilarity of complex networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, 71(016127), January 2005.
- L. Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18:39–43, 1953.
- P. J. Mucha, T. Richardson, K. Macon, M. Porter, and J-P. Onnela. Community structure in Time-Dependent, multiscale, and multiplex networks. *Science*, 328(5980): 876–878, May 2010. ISSN 1095-9203. doi: 10.1126/science.1184819. URL <http://dx.doi.org/10.1126/science.1184819>.

- Z. V. Stoyanov, G. Smith, B. Chakrabarti, P. Grindrod, D. V. Greetham, D. Saddy, T. Lee, T. Johnstone, and J. Gledhill. High resolution communicability measures of performance across fmri data from human brains. 2013. To be submitted.
- V. Tozzini. Coarse-grained models for proteins. *Curr. Opin. Struct. Biol.*, 15(2):144–150, April 2005.