Graph coarse-graining reveals differences in the module-level structure of functional brain networks

Rainer Kujala, Enrico Glerean, Raj Kumar Pan, Iiro P. Jääskeläinen, Mikko Sams and Jari Saramäki
1Department of Computer Science, Aalto University, PO Box 15400, FI-00076 Aalto, Finland
2Department of Neuroscience and Biomedical Engineering, Aalto University, Aalto, Finland

Keywords: community detection, functional connectivity, functional magnetic resonance imaging, modularity

Abstract

Networks have become a standard tool for analyzing functional magnetic resonance imaging (fMRI) data. In this approach, brain areas and their functional connections are mapped to the nodes and links of a network. Even though this mapping reduces the complexity of the underlying data, it remains challenging to understand the structure of the resulting networks due to the large number of nodes and links. One solution is to partition networks into modules and then investigate the modules’ composition and relationship with brain functioning. While this approach works well for single networks, understanding differences between two networks by comparing their partitions is difficult and alternative approaches are thus necessary. To this end, we present a coarse-graining framework that uses a single set of data-driven modules as a frame of reference, enabling one to zoom out from the node- and link-level details. As a result, differences in the module-level connectivity can be understood in a transparent, statistically verifiable manner. We demonstrate the feasibility of the method by applying it to networks constructed from fMRI data recorded from 13 healthy subjects during rest and movie viewing. While independently partitioning the rest and movie networks is shown to yield little insight, the coarse-graining framework enables one to pinpoint differences in the module-level structure, such as the increased number of intra-module links within the visual cortex during movie viewing. In addition to quantifying differences due to external stimuli, the approach could also be applied in clinical settings, such as comparing patients with healthy controls.

Introduction

Functional magnetic resonance imaging (fMRI) data comprising blood-oxygen-level-dependent (BOLD) time series across the brain are increasingly being analyzed with the tools of network science (Bullmore & Sporns, 2009; Newman, 2010; Power et al., 2011; Papo et al., 2014). In this approach, the functional organization of the brain is modeled as a network, or a graph, where nodes correspond to different brain areas and links indicate dependencies between the BOLD time series of the brain areas. Whenever the dependency between two BOLD signals is strong enough, the respective brain areas are thought to be functionally related. Because of this, one is typically only interested in links for which the measured dependency is strong, and only those links are included in the network describing the functional organization of the brain.

Network analysis has revealed many insights on the functional structure of the brain. On the scale of network nodes, the use of various network centrality measures has allowed consistent detection of certain hub regions (van den Heuvel & Sporns, 2013). At the network level, the entire structure of functional brain networks has been found to be of the ‘small-world’ type (Watts & Strogatz, 1998; Eguiluz et al., 2005; Salvador et al., 2005), meaning that the average number of steps required to reach any node in a network is low while the network is locally clustered. In addition to revealing such general properties, the network framework has also been used for investigating how brain dynamics depend on different stimuli (Lahnakoski et al., 2012a), mental health (Achard et al., 2012; Alexander-Bloch et al., 2012; Glerean et al., 2016) and age (Meunier et al., 2009a).

Despite being widely applied, the network approach to brain functional networks can still be considered as somewhat immature. There is no single commonly accepted way of constructing and analyzing networks, and methodological variations persist in the literature (Stanley et al., 2013; Garrison et al., 2015). In particular, there is no standard set of brain areas to be used as network nodes; rather, different studies use different definitions of what constitutes a node (Stanley et al., 2013). Approaches for defining network nodes range from the use of anatomical atlas-based Region-Of-Interest (ROI) parcellations to the use of the original fMRI imaging voxels of a few mm³ in size (Stanley et al., 2013). Typically, the number of nodes in atlas-based definition schemes is of the order of 10², while the number of nodes in voxel-based node definition schemes is typically of the order of 10⁴–10⁵. These differences between node definitions make comparison of results across studies challenging.
When atlas-based ROI parcellations are used, the resulting small number of nodes allows meaningful network visualization and makes network analysis easier. However, the low resolution of this scheme also means that the nodes may cover multiple functionally specific areas. The BOLD signal representative for each node is typically computed as an average over the node’s spatial extent, which can lead to significant loss of information. In the worst case, if the area covered by the node overlaps with multiple functional areas, the averaged signal may not be representative of any true underlying brain function. Therefore, if the maximum possible amount of information is to be retained, the recommended way is to use individual voxels as the network nodes (Stanley et al., 2013), i.e. to work with far larger numbers of nodes. Given that with high-field (e.g. 7 Tesla) fMRI it has been possible to reach a spatial resolution smaller than 1 mm³ (Yacoub et al., 2008), the volumes that voxel-based nodes represent can be expected to decrease in the future, resulting in even more nodes.

The downside of the increased level of detail of the voxels-as-nodes approach is that the large numbers of nodes and links make network analysis more challenging. The problems are both technical (e.g. it is difficult to deal with correlation matrices of $10^6 \times 10^6$ elements) and methodological. While one can fairly easily compute node-level measures such as centralities and global network characteristics from degree distributions to average path lengths, understanding the overall organization of a network’s links becomes difficult. This is related to the scales where the interesting features often live – beyond the level of individual nodes but below the network level. Because of this, e.g. comparing two voxel-level networks on a link-to-link basis becomes meaningless. Further, while network visualization is typically used for an overview of network structure, it only works well for smaller networks.

One typical approach to detecting structural features on the intermediate level between nodes and the network is to split the network into modules. Loosely speaking, these are groups of nodes that are densely connected internally but have sparse between-group connections (Fortunato, 2010). There are many more detailed definitions of modules, and accordingly, many different algorithms for detecting modules in networks. One common way is to use stochastic algorithms that try to optimize some quality measure for partitioning the network into non-overlapping groups of nodes. For examples of applications of module detection to functional brain network analysis, see (Meunier et al., 2009b, 2010; Power et al., 2011; Alexander-Bloch et al., 2012; Uehara et al., 2012; Betzel et al., 2016; Glerean et al., 2016; Sporns & Betzel, 2016).

Any observed modular structure of functional brain networks may arise from experimental conditions (different stimuli), or reflect more persistent, underlying features of the subject’s brain. Some recent studies have investigated how network modules differ between healthy controls and patients suffering from schizophrenia (Alexander-Bloch et al., 2012), patients with autism spectrum disorders (Glerean et al., 2016) or patients who are comatose (Achard et al., 2012). However, because of various intricacies in the detection of network modules and the difficulty of assessing differences between module partitions, results have not been straightforward to interpret in terms of brain function, or the statistical significance of findings has remained elusive. Although there are statistical approaches to assess the stability of modules (Moussa et al., 2012; Glerean et al., 2016), or the similarity of partitions between groups of networks (Alexander-Bloch et al., 2012), there are no standard ways for verifying the statistical significance of specific differences between two partitions. This is a serious limitation for the module detection approach because it is often the specific differences that are most relevant – e.g. if there is a module in one network that is split into two parts in another, does this carry meaningful information or is it purely due to chance?

Therefore, there is a need for appropriate methods for analyzing and comparing functional network structure at the intermediate level of modules. We argue that instead of focusing on how the network modules themselves differ between groups of networks, it is more fruitful and sound to assess the differences in the numbers of connections between and within a fixed set of modules that is used as a common frame of reference for all groups. The boundaries of these modules are defined using a specific network or group of networks, with a module-detection method of choice, and applied as such to the rest of available networks. Then, one focuses on differences in connectivity within and between the reference modules. This coarse-graining approach allows transparent, statistically verifiable investigation of module-level differences between networks.

Below, we first demonstrate the difficulties of comparing network modules with the help of a toy example, and proceed to show how the coarse-graining approach overcomes these problems. Then, to show the applicability of our approach in practice, we apply it to networks constructed from fMRI data recorded for 13 subjects during rest and movie viewing.

Comparing the structure of functional networks – coarse-graining and alternative approaches

The simplest way of assessing differences in the link structure of functional brain networks is to investigate how the existence or weight of individual links differs between groups of networks corresponding to, e.g. different experimental conditions or subject populations. In this case, assessing the statistical significance of the observed differences is straightforward. However, with voxel-based functional brain networks that have thousands of nodes and millions of links, this approach is impractical for several reasons. First, the number of statistical tests that need to be performed reaches millions, requiring a large amount of computational resources. Second, even when best efforts are made to standardize the presentation of brain imaging data across subjects by transforming it into standard coordinates, the functional correspondence between specific nodes and links across subjects is never perfect, due to slight differences in the anatomy and the location of functional regions between individuals. Third, structures of interest may span large numbers of nodes and links, which is not captured by comparing these in isolation. Finally, visualizing all differences in links between groups of networks becomes challenging due to the large number of possible links.

Because of the above, some studies have focused on quantifying differences between networks at the level of modules that have been discovered using stochastic network partitioning algorithms (Alexander-Bloch et al., 2012; Glerean et al., 2016). However, there is no universally agreed definition for network modules (Fortunato, 2010). Instead, almost every algorithm introduces its own definition of a network module that dictates how it partitions a network and how modules should look like. Subsequently, different algorithms give different results. Therefore, proper interpretation of module partitions requires a profound understanding of the underlying definition of a module and of the actual implementation of the algorithm.

To ensure the statistical significance of the discovered modules, it is possible to measure the significance of a discovered module with respect to a random network null model (Lancichinetti et al., 2010, 2011). There are also methods for measuring the overall level of similarity between two distinct partitions, including the Variation of

© 2016 Federation of European Neuroscience Societies and John Wiley & Sons Ltd
European Journal of Neuroscience, 44, 2673–2684
Information criterion and the adjusted Rand index (Hubert & Arabie, 1985; Meilă, 2007). For a review of the variety of available methods, see Meilă (2007).

Recently, methods for comparing similarity of partitions have also been applied to measure differences between partitions of functional brain networks of two subject populations (Alexander-Bloch et al., 2012; Glerean et al., 2016). However, in particular when applied to groups of networks, the power of such methods is limited – they can only tell whether there is a significant difference between partitions or not. Then, if there is a difference, one still needs to find out in ad hoc ways where the difference comes from. Further, the lack of overall significant differences may still hide smaller scale systematic differences.

When working with two groups of partitions, for an overview of their differences, it is also common to define representative partitions for each of the conditions (Alexander-Bloch et al., 2012; Glerean et al., 2016); these may allow visual comparisons. In any case, any observed differences between the representative partitions have to be validated. Yet, quantifying the statistical significance of any specific difference between the partitions is challenging. This difficulty is exemplified in the following: consider that there are two experimental conditions, 1 and 2, and there is a representative partition for both conditions. Consider then that module A of condition 1 is split into modules B and C in condition 2. Is this split meaningful and consistent across individuals, or associated with noise and the stochastic nature of the module detection pipeline used? There is no simple and principled way to answer this question. Now add to the level of difficulty – there are other modules too in both conditions, and one can construct a mapping between these – perhaps modules D and E of condition 1 partially correspond to module F of condition 2 that also shares some nodes with module G of condition 1, and so forth (see Fig. 4). At the moment, there are no statistical frameworks for directly assessing the statistical significance of such differences. This is a serious problem because most module detection methods are stochastic, producing a set of outcomes instead of a single outcome, and can be rather sensitive to small variations in links or their weights. While investigating the consistency of module assignment within a group of networks may help (Alexander-Bloch et al., 2012; Glerean et al., 2016), this approach provides no solution to common practical problems, such as assessing whether the splitting of a specific module is real or possibly just due to chance.

Because of the above-mentioned sensitivity of module-detecting algorithms, it may even be difficult to disentangle statistical significance from unwanted effects arising from this sensitivity. This is because small differences in a network’s link structure may give rise to large differences in the module partition. These small differences may be significant and consistent – it is the size of the effect they cause on the modular structure that is problematic. At other times, large structural differences result in little or no differences in the discovered modules. We illustrate this with Fig. 1. In this toy example, we consider two networks, A and B, that have a clear modular structure. Both networks share the same set of nodes and their link structures are almost the same, with four links being wired differently in B. We then apply the Louvain algorithm (Blondel et al., 2008), one of the most popular methods for detecting network modules, to the two networks separately. While the removal of three links between the yellow and brown modules of Network A results in no differences in the modules detected in network B, the addition of only one extra link between the two smaller leftmost modules in Network A (green, violet) forces them to merge into one large blue module in Network B. This simple example shows that it is not straightforward to infer the underlying differences in the overall organization of links by comparing network partitions – the cause and the effect size can be disproportionate.

When dealing with noisy experimental fMRI data, the challenges are even greater. The randomness in module partitions that arises from the stochasticity of the algorithms is further amplified by the

![Diagram of networks and their coarse-grained versions](image)

**Fig. 1.** Network modules, the difficulty of comparing them, and the strength of network coarse-graining. **Top row** shows two networks A (left) and B (middle) that differ from each other only by relocation of four links (right); blue links are present in A but missing in B, whereas red links are there in B but not in A. **Second row** shows the network modules corresponding to networks A and B as identified by the Louvain algorithm that optimizes modularity (Blondel et al., 2008; Newman & Girvan, 2004). On the right, the differences in the discovered modular structure are visualized as an alluvial diagram (Rosvall & Bergstrom, 2010) – the left side of the diagram represents the modules of network A and the right side represents the modules of network B. Ribbons connecting the left and right sides show how the modules of A and B match each other in terms of node composition – here the only difference between A and B is that the green and violet modules of A are subsets of the blue module in network B. This difference arises from the addition of a single link between the green and violet modules. To the contrary, the three links between the yellow and brown modules in network A that are relocated in B do not give rise to differences in the modules; the rest of changes (see top row, right) do not produce any changes in modules either. These examples highlight the difficulty in inferring differences in the link structure of networks based on the modular structure alone. **Third row** shows coarse-grained versions of networks A and B and their difference, the coarse-graining being based on the modules detected for A. Here, the width of the link between two modules corresponds to the number of links between their constituent nodes in the original network. Similarly, the width of the arc around each module represents the number of links within the module. The differences between the coarse-grained networks are shown on the right. For the blue links, there are more between-module links in A than in B, and for the red links, B has more between-module links. Note, how the coarse-grained difference network is able to compactly summarize the differences between networks A and B. **Bottom row** shows the same information as the third row, but in the matrix form. Each (square) element of the matrix corresponds to a module, and the area of the square is proportional to the number of links between modules (off-diagonal) or within modules (diagonal). The row and column colors correspond to the modules of network A.
noise, which obscures real structural differences between networks. In addition, there are no statistical frameworks for directly assessing the statistical significance of specific differences between partitions. Thus, even when the structure of network modules sheds some light on how the brain functions under different conditions, drawing definite conclusions from the differences between network partitions is difficult.

If direct comparison of module partitions as well as individual links is difficult, is there any way around the problem? We argue that combining both points of view yields fruitful results. In particular, structural differences between networks are revealed more clearly by first producing a fixed set of modules to be used as a frame of reference, and then comparing the strength of the inter- and intra-module connections across networks using the same module boundaries. This approach circumvents the difficulty of directly matching and comparing modules, and at the same time retains information carried by individual links. As an example, if network X has a strongly modular structure and its modules are used as the frame of reference, it will have dense within-module connections and sparse between-module connections. If network Y has no modular structure, it will have much sparser within-module connections and much denser between-module connections than X in this frame of reference.

The above approach corresponds conceptually to the network coarse-graining, where each module corresponds to a node of the coarse-grained network, and the number of links between two modules in the original network corresponds to the weight of the link in the coarse-grained network. The number of internal links within each module is taken into account as the weight of the self-link (connection from the node to itself) of each module-node of the coarse-grained network. In principle, the modules can be determined by applying any module detection method of choice, or by using network partitions independent of data (e.g. reported by some previous study) – their choice only affects the clarity of results, not the coarse-graining approach that follows or the statistical significance of findings.

We demonstrate the usefulness of the coarse-graining approach with the same toy networks as earlier in Fig. 1. Notably, the comparison of coarse-grained networks reveals differences in a more transparent way than when attempting to compare modules. Beyond this simple example, when groups of networks are to be compared, the coarse-graining approach allows straightforward statistical testing of the differences in the mean number of links within a module or between two modules.

Materials and methods

Participants

The participants were 13 healthy native Finnish speakers (ages 22–43 years, 2 females, 2 left-handed, no neurological or psychiatric history, no hearing impairments, normal vision). The ethical committee of the Hospital district of Helsinki and Uusimaa granted permission for this study which was conducted in accordance with the guidelines of the declaration of Helsinki. Each subject gave written informed consent prior to participation.

Stimulus paradigm

The stimulus used in this paper has also been used previously (Lahnakoski et al., 2012b; Salmi et al., 2014) and consisted of an edited version of the Finnish movie ‘The Match Factory Girl’ (directed by Aki Kaurismäki, 1990). The film was projected on a semi-transparent screen behind the subject’s head and the audio track was delivered via plastic tubes through porous earplugs. Each subject went through three sessions with the following order: resting state (15 min, 450 volumes), free viewing of the film (22 min 58 s, 689 volumes), resting state (15 min, 450 volumes). In this study, we analyze the data recorded during rest before viewing the movie, and during movie viewing. After preprocessing, the movie session was truncated to match the length of the rest session to avoid any biases from different scan durations.

Data acquisition

Magnetic resonance imaging (MRI) was conducted on a 3.0 Tesla GE Signa Excite MRI scanner, with a quadrature 8-channel head coil. For each volume, a total of 29 functional gradient-echo planar axial slices (thickness 4 mm, 1 mm gap between slices, in-plane resolution 3.4 mm × 3.4 mm, imaging matrix 64 × 64, echo time (TE) 32 ms, repetition time (TR) 2000 ms, flip angle 90°) were collected. T1-weighted images were also acquired (TE 1.9 ms, TR 9 ms, flip angle 15°; SPGR pulse sequence) with in-plane resolution of 1 mm × 1 mm, matrix size 256 × 256 and slice thickness 1 mm with no gap.

Preprocessing

Preprocessing of the fMRI data was carried out with FSL (release 4.1.6, www.fmrib.ox.ac.uk/fsl). The first 10 volumes of each session were discarded from the analysis. Motion correction was performed with McFlirt and the data were spatially smoothed using a Gaussian kernel with 6 mm full-width half maximum and high-pass filtered with a 100 s cutoff. Functional data were co-registered with FLIRT to the anatomical image allowing 7 degrees of freedom. Furthermore, the data were registered from the anatomical space to the MNI152 2 mm standard template (Montreal Neurological Institute), allowing 12 degrees of freedom. The signal was bandpass-filtered with a passband of 0.01–0.08 Hz in accordance with standard functional connectivity procedures. To control for motion artifacts, motion parameters were regressed out from the data with linear regression (36 Volterra expansion based signals, see Ref. (Power et al., 2014)). As it is known that head motion affects connectivity results, we controlled for motion with framewise displacement (Power et al., 2012) – all subjects had at least 95% of time points under the suggested displacement threshold of 0.5 mm. For this reason, we decided not to use the scrubbing technique and utilize all time points. When looking at the individual mean framewise displacement, there was no significant difference between conditions (P = 0.2732). Finally, to further control for artifacts, voxels at the edge between brain and skull where the signal power was < 2% of the individual subject’s mean signal power were excluded from the analysis. This resulted in 5562 6-mm isotropic voxels of brain gray matter covering the whole cerebral cortex, subcortex and cerebellum. After removing the first and last 15 data points due to bandpass filtering artifacts (Power et al., 2014), we obtained for each 6-mm voxel a BOLD time series with 410 time points corresponding to a duration of 13 min 40 s.

Network construction

The functional dependency of two voxels i and j can be measured in many ways, given their BOLD time series sj(t) and sj(t) (Smith et al., 2011). As there is no consensus on the best measure, we opt for simplicity and use the Pearson correlation coefficient, which has

© 2016 Federation of European Neuroscience Societies and John Wiley & Sons Ltd

European Journal of Neuroscience, 44, 2673–2684
been shown to capture a major proportion of pairwise dependencies in fMRI data (Hlinka et al., 2011). For each subject and condition, we then compute a correlation matrix $\mathbf{R}$, whose elements $r_{ij}$ are the estimated Pearson correlation coefficients between the time series of voxels $i$ and $j$. Given that we have two different conditions and 13 subjects, this yields 26 correlation matrices in total.

In principle, it would be possible to treat the full correlation matrices as weighted networks and directly subject them to analysis. However, to avoid computational problems from the large number of entries, and to get rid of low-valued entries that clearly do not indicate functional connections, we take the common approach of thresholding the matrices so that only the strongest links remain, and treat the resulting sparse matrices as unweighted networks.

There are multiple possible ways of thresholding the matrices. One common approach is to use a constant threshold value, so that only nodes pairs whose correlation coefficients exceed this value are connected by a link. Another typical approach is to only include a fixed fraction of the strongest links in the network. Here, we adopt the latter approach as it has been shown to provide more stable estimates of various network measures (Garrison et al., 2015), and because comparing networks that have the same number of links is more transparent.

We construct networks from correlation matrices as in (Alexander-Bloch et al., 2010, 2012; Glerean et al., 2016) – for each matrix, we begin with an empty network, where each node corresponds to a voxel. We then first compute the maximum spanning tree (MST) of the correlation matrix, and insert the corresponding links to the network. As the MST connects all nodes, this guarantees that no nodes or groups of nodes remain isolated in the network; isolated modules would cause technical difficulties in the later stages of our pipeline. Next, we sort all correlation coefficients, and insert links corresponding to the strongest positive coefficients until the network contains $\frac{1}{2}N(N-1)\rho$ links in total, where $\rho \in [0, 1]$ is a predefined network density. As the end result, we obtain 26 undirected, unweighted networks that all share the same set of nodes, and have the same number of links.

**Selection of network density**

Choosing the fixed density value for thresholding the networks is not straightforward. There are no commonly accepted criteria for an optimal density. If the density is very low, too much information is discarded and features of interest may be removed. On the other hand, if there are too many links, their presence may obscure relevant structures. Sometimes this problem can be overcome by investigating network structure across different network densities (Alexander-Bloch et al., 2012; Lord et al., 2012). However, carrying out detailed structural analysis of a large number networks of different densities quickly becomes overwhelming. Then, selecting a reasonable, specific network density may be a better option. This is also the case in this study, where we compare the module-level structure of groups of networks in detail.

To guide our choice of network density, we apply a recently proposed diagnostic for selecting network density, first adopted in (Glerean et al., 2016), and investigate the similarity of pairs of networks as a function of the network density. If two networks share the same set of nodes and have the same number of links, the most straightforward approach to measure their similarity is to count the number of common links. This is the approach we adopt. Given two networks $G$ and $G'$ that both have $\frac{1}{2}N(N-1)\rho$ links, we monitor how the fraction of links common to the networks $f(G, G')$ changes with the network density $\rho$. The fraction of common links $f(G, G')$ is defined as the number of shared links divided by the total number of links in one network:

$$f(G, G') = \frac{\text{Number of common links in } G \text{ and } G'}{\frac{1}{2}N(N-1)\rho} \in [0, 1]. \tag{1}$$

In Fig. 2, we show $f$ as a function of the network density when averaged over pairs of networks, so that each pair represents (i) the same subject in different conditions, (ii) different subjects in the rest condition, (iii) different subjects in the movie condition and (iv) different subjects in different conditions. To illustrate the dependency of $f$ on network density, we also show a reference curve corresponding to the expectation value of a null model, where the links are assigned between random node pairs. In the null model, the mean fraction of common links equals $\rho$, and thus $f$ grows linearly with density.

For all cases, the fraction of common links first increases until $\rho \approx 0.1\%$; one might envision that at this density, a common ‘backbone’ shared by networks is well captured. Then, $f$ decreases until $\rho \approx 2\%$, after which it begins to monotonously increase as the networks become denser and more and more links are necessarily shared, as also indicated by the reference curve. Note that before $\rho \approx 2\%$, the experimental values of $f$ are much above the reference, which indicates that the observed dependency between $\rho$ and $f$ cannot be explained by random chance alone; this directly means that the network structure also deviates from random.

There are multiple ways of choosing the threshold density based on the diagnostic. One possibility is to select the density with maximal difference to the reference, maximizing the non-trivial similarity between network pairs. Another option is to pick values around the
dip of $f$ from 1% to 10%, in order to maximize variation between network pairs while retaining more similarity than for very low values of $f$. We adopt the latter criteria, and pick the value $\rho = 2\%$ to be used in all subsequent analyses. This choice is also compatible with previous studies, where the same network density has been used (Alexander-Bloch et al., 2012, 2013; Glerean et al., 2016). In our data, the 2% network density translates to 309 303 links in each thresholded network; for different networks, this corresponds to a correlation coefficient threshold of 0.56 ± 0.04 (std) on average (no statistically significant difference between conditions).

Figure 2 also reveals some further insights. First, the network similarity $f(G, G')$ is remarkably higher for networks corresponding to the same subject in different conditions than for different subjects in the same condition. This indicates that individual variation dominates over differences caused by different stimuli. Subsequently, it is essential to take the paired nature of the data into account when validating any results statistically. Second, when the similarity of networks of different subjects is assessed, network pairs corresponding to the same condition are seen to be more similar than network pairs corresponding to rest condition or different conditions. This is expected, as the viewing of a well-directed movie stimulus has been found to synchronize the subjects’ brains (Hasson et al., 2010), which results in increased functional connectivity compared to the similarities of resting state networks arising from shared functional anatomy and connectivity.

**Computation of condition-wise consensus partitions**

Given the networks for all subjects S1–S13 and conditions, our next target is to compute representative network partitions for each condition. As outlined in Fig. 3, this is realized in two steps: First, we partition the networks using a popular partitioning algorithm, the Louvain algorithm (Blondel et al., 2008), which is based on modularity optimization (Newman & Girvan, 2004). In more detail, we run the stochastic Louvain algorithm 100 times for each network and select the network partition with the highest value of modularity. Then, for each condition, we summarize all 13 selected partitions with the help of the MCLA meta-clustering algorithm (Stuehl & Ghosh, 2003) which yields one representative network partition as an output. Internally, MCLA first constructs a network of all the modules of the input partitions, where the weights of the links between the modules reflect the similarity of the modules as measured by the Jaccard index. This module-network is then partitioned into a set of ‘meta-clusters’ using the network partitioning algorithm METIS (Karypis & Kumar, 1998), which takes the number of modules as a parameter. In this study, we have set the number of modules to be found as the median number of modules present in the input partitions. The discovered meta-clusters then provide the basis for the actual consensus modules – each node is assigned to the meta-cluster in which it is most consistently present, while a consistency value is normalized by the total number of modules in a meta-cluster.

To verify the representativeness of each of the consensus modules, we compute the distances between the original partitions, and compare this distribution to the similarities between the consensus partition and the individual subject-wise partitions. The specific similarity measures that we use are normalized mutual information (Danon et al., 2005), the adjusted Rand index (Hubert & Arabie, 1985) and the variation of information (Meilă, 2007), which is a distance measure between partitions.

In summary, the above pipeline thus transforms the subject-specific networks corresponding to one condition to a single representative consensus partition $\mathcal{P}$ consisting of $m$ modules $C_1, \ldots, C_m$ which in turn are sets of network nodes such that each node belongs to exactly one module.

Note that both the discovery of the subject-wise network partitions as well as the computation of the representative network partitions can be done in many different ways, without affecting how the core of the proposed pipeline, the coarse-graining and subsequent network comparison, are carried out. For detecting modules, there are many other options than the Louvain method used here, and even the Louvain method itself can be parametrized differently to emphasize smaller or larger modules with the resolution parameter $\gamma$ (Reichardt & Bornholdt, 2006; Fortunato & Barthelemy, 2007). In this study, we have opted for simplicity and used the default $\gamma = 1$ as this choice already yielded a meaningful number of modules for summarizing differences between networks.

Similarly, the consensus partitions can also be computed using a different algorithm (see e.g. Ref. (Lancichinetti & Fortunato, 2012). Further options include setting a different number of modules to be found by METIS, and using all the 1300 discovered partitions as inputs to the MCLA algorithm as it is known that modularity can have multiple high-scoring local maxima corresponding to very different network partitions (Good et al., 2010). Moreover, instead of computing consensus partitions, one could also choose the condition-wise representative network partitions to be the individually discovered network partition whose distance to other partitions as measured, e.g. by normalized mutual information or the adjusted Rand index is the smallest. Again, we have opted for simplicity, and used the MCLA approach as it has been used previously in the literature (Glerean et al., 2016).

**Comparing groups of coarse-grained networks**

The network coarse-graining process briefly introduced in Fig. 1 is defined in detail as follows: Given a network $G$ and a partition $\mathcal{P}$ consisting of a set of modules $C_1, C_2, \ldots, C_m$, the coarse-graining process yields a matrix $W$ that has the following properties: The non-diagonal matrix elements $W_{ij}$ represent the total number of links between the nodes (voxels) belonging to $C_i$ and $C_j$. Similarly, each diagonal element $W_{ii}$ represents the number of links within module $C_i$.

To evaluate differences between experimental conditions, we first coarse-grain each subject’s network into its matrix representation $W_{ij}$, where $i$ stands for the index of the subject. Then, we average the coarse-grained matrix representations of all 13 subjects over each condition:
\[ \langle W \rangle_{\text{rest}} = \frac{1}{13} \sum_{i=1}^{13} W_{\text{rest}, i}, \] (2)

and

\[ \langle W \rangle_{\text{movie}} = \frac{1}{13} \sum_{i=1}^{13} W_{\text{movie}, i}. \] (3)

Then, by investigating the elements of the mean difference matrix \( \Delta W = \langle W \rangle_{\text{movie}} - \langle W \rangle_{\text{rest}} \), we can quantify the level of differences in the numbers of connections between and within modules. Thus, the value of \( \Delta W_{i,j} \) indicates how many more connections there are on average between modules \( i \) and \( j \) in the movie condition than in the rest condition.

To test the statistical significance of our findings, we perform paired permutation tests separately on each of the matrix elements. As we have a limited number of subjects, we use the full permutation distribution available, yielding in total \( 2^{13} = 8192 \) different permutations of the 13 subjects. All P-values we report are two-sided, and we correct for multiple comparisons using the original Benjamini–Hochberg (BH) FDR correction (Benjamini & Hochberg, 1995).

**Code**

Python code for the coarse-graining pipeline is available as a part of the brainnets library for analyzing fMRI data http://github.com/rmkuhal/brainnets. The library is accompanied by an example for running the pipeline, and it has options for, e.g. changing the method used for module detection.

**Results**

**Consensus modules are similar to previously reported resting-state modules**

The consensus modules computed for the rest and movie-viewing conditions are shown in Fig. 4A with different colors on the cortical surface. In the Supporting Information, we report the number of modules in the original partitions, and demonstrate that the consensus modules are representative of the original condition-wise partitions with similarity comparisons. A browsable display of each module is available at NeuroVault http://neurovault.org/collections/1080/ (Gorgolewski et al., 2015), where the modules are weighted by their consistency as measured by scaled inclusivity (Steen et al., 2011) (see Data S1 for more details). In the center of Fig. 4A, the matching of the rest and movie consensus partitions is also visualized as an alluvial diagram (Rosvall & Bergstrom, 2010). Statistics on the number of modules in the subject-wise partitions, and verification of the representativeness of the consensus partitions are presented in Data S3.

For the rest condition, we identified 11 consensus modules (Fig. 4, left-hand side of alluvial diagram, from bottom to top): (i) Limbic (LIM) subcortical midbrain structures; (ii) Cerebellum/ventro-temporal (CRBL/VT); (iii) Default mode (DM); (iv) Precuneus (PCUN); (v) Visual (VIS); (vi) Auditory (AUD); (vii) Salience (SAL); (viii) Fronto-parietal (FP); (ix) Dorsal attention (DA); (x) Sensorimotor (SM). For details on how the labels were assigned to modules, please see Data S1.

There is a good agreement in the literature on the module-level structure of resting-state networks, which have been identified using various methods such as multidimensional clustering (Yeo et al., 2011), Infomap graph clustering (Power et al., 2011), and independent component analysis (Smith et al., 2009). Overall, the resting-state consensus modules we obtained are in line with the previously reported resting-state modules (see Data S1).

However, there is no general agreement on the module structure during movie viewing, or more generally, during a task. While one study has found that task and rest are highly similar (Cole et al., 2014), another study has found remarkable differences in subcortical, limbic regions as well as primary sensory and motor cortices (Mennes et al., 2013). In our case, overall, the movie consensus modules are similar to the resting-state modules, and e.g. the dorsal attention module that has been previously reported in resting-state studies (Yeo et al., 2011; Cole et al., 2014) is even better identified in our movie consensus partition than in the rest consensus partition (see Data S1). Interestingly, we also identified a VTL subnetwork present only during movie viewing in agreement with Glerean et al. (2016), possibly suggesting stronger functional couplings between the brain areas involved in the processing of social and emotional events in the movie.

**Differences in condition-wise consensus modules are difficult to interpret**

The main motivation behind our coarse-graining approach is that direct comparison of modules is rather difficult and the results are too hard to interpret. We demonstrate this with the consensus modules of both conditions. As shown in Fig. 4A, the consensus modules obtained for the movie and rest conditions are broadly speaking similar – for most rest modules, there is a clear counterpart among the movie modules. At the same time, almost all rest modules overlap with multiple movie modules (and vice versa). There are no simple relationships such as one module splitting into two, and the varying amount of overlap between modules results in a diagram that is not straightforward to interpret.

As there are no statistical frameworks that can be used for measuring the significance of the relationships between multiple modules, the differences in the alluvial diagram lack statistical validation. Some insights into the significance of the transitions in the alluvial diagram could be obtained by investigating the consistency of the modules (Moussa et al., 2012; Glerean et al., 2016). However, these methods do not directly assess the significance of individual splits and merges of modules between partitions. Thus, it is challenging to draw conclusions on structural differences between the underlying networks based on the modules and the alluvial diagram alone.

**Network coarse-graining provides results that are straightforward to interpret and statistically verifiable**

In Fig 4B, we show the matrix representations of the average coarse-grained networks corresponding to both conditions, as well as the coarse-grained difference network, where the resting-state consensus modules have been used as the basis for coarse-graining. As expected, the matrices representing the coarse-grained networks have high values on their diagonals, indicating that the density of links within consensus modules is higher than between them. Off-diagonal elements provide an overview of how strongly the modules are connected.
Fig. 4. Panel A – Differences in consensus modules are difficult to interpret. On the left and right, we show the consensus modules obtained for all networks corresponding to the rest and movie conditions, respectively. The alluvial diagram displayed between the modules of different conditions shows how they relate to one another in terms of node membership. The height of each module and of the ribbon connecting two modules is proportional to the associated number of nodes. The colors have been chosen by maximal match between rest and movie modules, to allow visual comparison. The abbreviation of the module’s label is shown next to it (see text for details). Further, each ribbon has also been labeled based the anatomical brain areas that the nodes constituting the ribbon belong to. For details on the labeling of the modules and ribbons, see Data S2. Panel B – Coarse-grained networks enable transparent detection of differences. In the three different plots, we show the average coarse-grained networks obtained using the consensus rest modules. On the left, the average coarse-grained networks for both rest and movie conditions are shown. The size of the black rectangles indicate the number of links between modules (within modules for diagonal rectangles). Note that the matrices are symmetric by construction, and thus the upper and lower triangles of the matrices contain the same information. On the right, the average coarse-grained difference matrix is shown. Note that the scale for square sizes is different from the left. The color of each matrix element in the difference matrix indicates the (uncorrected) $P$-value obtained from a mean difference permutation test. The size and color of a square thus indicate both the associated effect size and the statistical significance of the difference. The color bar shows the 0.05 FDR threshold computed with the BH procedure. In total, there are five differences that survive the FDR correction; each of those indicates more connections in the movie condition between certain modules.
In the coarse-grained difference matrix, we observe multiple elements that survive the 0.05 Benjamini–Hochberg FDR correction. These are also listed in Table 1. All surviving elements are positive, indicating that there are more connections between the modules in the movie networks. In particular, the VIS module displays significantly more external as well as internal connections in the movie condition. The number of connections between the AUD and DM modules is also increased in the movie condition. The coarse-graining method thus succeeds in highlighting task-driven changes at visual areas as well as inferior temporal structures. For similar coarse-graining results where the movie modules are used as the frame of reference, please see the Supporting Information.

There are of course some similarities between the coarse-grained difference network and the alluvial diagram presented in Fig. 4. As an example, in both it is seen that the resting-condition VIS module and part of the resting-condition VISx module merge to form the larger VIS module in the movie condition. A simple explanation for this would be that there are more connections between the VIS and VISx rest modules in the movie condition. However, as discussed in section ‘Comparing the structure of functional networks – coarse-graining and alternative approaches’, there are other possible explanations, including the high sensitivity of the module-detection algorithms to small changes. Therefore, this explanation should be statistically validated. This is not possible when directly comparing the modules, and the alluvial diagram provides no help. However, the coarse-grained difference matrix clearly indicates that there is a statistically significant increase (47%) in the number of links between the VIS and VISx modules in the movie condition. Thus, while the alluvial diagram can be used for formulating hypotheses on changes in network structure, the coarse-graining process allows verifying that the observed differences are not due to random chance.

### Discussion

We have introduced an approach that allows comparing the intermediate-level structure of functional networks, circumventing the problems of directly attempting to match module partitions of different networks. This approach is based on using a single set of modules as a frame of reference, and coarse-graining voxel-level networks according to this frame. The benefits of this approach are that the results are straightforward to interpret and can be verified statistically. As a testbed for this approach, we compared networks constructed from fMRI data recorded from 13 subjects during movie viewing and rest, and detected clear differences that were not seen when directly comparing modules. This demonstrates the strength of the coarse-graining approach in detecting intermediate-scale differences between functional brain networks; instead of differences due to external stimuli, the approach could also be applied in clinical settings for, e.g., comparing patients with healthy subjects.

### Alternatives to network coarse-graining

At first sight, network coarse-graining may seem similar to the traditional approach of first averaging the BOLD signal over all the voxels comprising a module or a ROI, and then inferring differences between conditions by comparing the correlations between the averaged time series of the modules. The important difference is that in the coarse-graining approach modules represent sets of voxels, and all information contained in the time series of their constituent voxels is retained. To the contrary, in the more traditional BOLD-averaging approach, even though the averaging can be argued to reduce noise, it also results in great loss of information. It is known that the average signal may not be representative of the voxels within a module (Craddock et al., 2012; Stanley et al., 2013). In the worst case, the average signal may not be representative of the voxels within a module or a ROI at all. Thus, networks constructed in this way do not contain information on the internal dynamics within modules/ROIs, unlike the voxel-level networks of the coarse-graining approach. As an example of this, with the BOLD-averaging approach, it would not have been possible to see that there are increased internal connections in the VIS module in the movie-viewing condition. The strength of the coarse-graining approach is that it retains all information on dynamics inside modules and helps to summarize it efficiently. Nevertheless, a more detailed comparison would be worth a further study.

In this study, we have applied coarse-graining to undirected and unweighted networks obtained with pruning the correlation matrices. While this process puts focus on the most meaningful links in the network, it may also result in the loss of useful information. However, the approach of assessing differences in inter-module and intra-module connectivity is general, and can easily be extended to weighted networks that take into account the strength of correlations between voxels, or directed networks that depict causal relationships between brain areas. If the full correlation matrix is used, network coarse-graining becomes similar to subnetwork analysis (Meskaldji et al., 2011) in terms of the underlying procedure, where one groups nodes based on a priori modules and then tests for differences in connection strengths between and within modules. It should be noted that while the main motivation for subnetwork analysis is to increase the power of statistical testing compared to link-wise statistical testing, the main motivation for network coarse-graining is to provide an easily interpretable way for understanding differences in the structure of functional networks. These two approaches thus arrive at a similar solution, although from different perspectives.

In a more general context of comparing of differences in weighted functional brain structures, network coarse-graining shares some similarity with the network-based statistic (NBS) (Zalesky et al., 2010) and spatial pairwise clustering (SPC) (Zalesky et al., 2012). These methods are both based on mass univariate statistical testing of each link in a graph. Then, links passing an a priori specified threshold are assigned to a graph, and grouped into link-clusters. NBS defines these link-clusters as the connected components of the graph, while SPC defines link-clusters such that links e and f belong to the same link-cluster if e’s incident nodes are spatial neighbors of f’s incident nodes. Then, a P-value can be obtained for the size of the clusters revealed by NBS and SPC by computing a reference distribution of link-cluster sizes by shuffling the end points of the supra-threshold links. The fact that NBS and SPC assume link-wise differences to form a connected component increases their statistical power, but simultaneously the assumption of connectedness can also restrict the possible features they can find. One such example is the case of “dispersed” differences in connections between modules that...
can be captured by network coarse-graining but not by NBS or SPC. On the other hand, some structures spotted by NBS or SPC might not be captured by coarse-graining. Thus, NBS and SPC are best regarded as complementary to network coarse-graining. However, combinations of NBS and coarse-graining are also possible – a large-link-cluster revealed by NBS can be efficiently summarized using network coarse-graining.

**Methodological considerations**

In our full analysis pipeline, there are several choices to be made before the actual comparison of networks. Already at the stage of data preprocessing, one has to choose whether or not one should regress out the global signal, or smooth the data spatially. In general, these questions are still open. In this study, we have chosen not to regress out the global signal; however, we have applied spatial smoothing. Note that we do not particularly advocate for either of these approaches. The use of spatial smoothing is expected to have some effect on the network structure, as it typically increases the number of short-distance links (even though neighboring voxels have a high probability to have a link between them even without smoothing (Salvador et al., 2005; Alexander-Bloch et al., 2013)). Sometimes short-distance links are excluded from the network in order to remove artifacts caused by smoothing or head movement (Power et al., 2011). However, we have not adopted this approach because neighboring nodes are likely to be functionally related, and thus such cutting out of links could result in a peculiar network structure that does not reflect the underlying brain function. In this study, we find a large fraction of long-distance links even after smoothing, and thus smoothing should not have biased our network structure too much (see Supporting Information). Given that there are few potential benefits in applying smoothing, and that the network coarse-graining method is tolerant to noisy link-level data, future studies can be performed without smoothing as well.

There are many choices to be made regarding the particulars of the module detection algorithm (and its parameters, if any). Naturally, these choices will also affect the exact outcome; however, whatever method is used to detect the modules, coarse-graining is a useful tool for comparing them. This is because the problems in interpreting differences between partitions are universal, e.g. for any module detection method, there is nearly always a borderline case where the existence or absence of one single link affects boundaries between modules. Naturally, the choice of a module detection method also affects the resulting coarse-grained networks; however, when comparing across conditions and using the modules of one condition as the frame of reference, the differences should still be straightforward to interpret. Finally, even though different module detection algorithms yield different network partitions, they are likely to capture some meaningful aspect of the network’s organization, which can then be passed on to the network coarse-graining stage.

**Selection of frame of reference**

The choice of the frame of reference is a key component in the application of network coarse-graining. This choice is not, however, as critical as it is in, e.g. when defining the ROIs in a ROI-level network, as the selection of the frame of reference does not affect the underlying model of the functional connectivity, i.e. the voxel-level network. In this study, this is demonstrated by the fact that the observed differences in the network structure remain relatively similar even if the movie modules were used as the frame of reference instead of the rest modules (see Supporting Information).

It is also worth pointing out that there is most likely no perfect partition of the brain into distinct functional areas. Instead of considering the multiplicity of possible partitions as a problem, it can rather be viewed as an opportunity – using different partitions for coarse-graining networks can actually turn out to be very useful. In the same way as one can deduce the shape of a three-dimensional object from its two-dimensional projections, a network’s structure can be better understood by investigating its different coarse-grained representations. Moreover, even though in this study we have used data-driven network modules as our frame of reference, one could equally well use a frame-of-reference based on some previous work. Examples include brain areas produced by Power et al. (2011), Yeo et al. (2011), and even any of the anatomical atlases, which would provide a frame of reference that is well known to researchers in the field. When comparing multiple networks corresponding to different tasks, it is perhaps best to use established brain parcellations for coarse-graining, as they provide a frame of reference which is easy to interpret.

**Conclusion**

We have developed a coarse-graining method to analyze differences in the modular structure of functional brain networks, and applied it to data recorded during rest and task. The coarse-graining approach focuses on differences in the connectivity between and within larger brain areas without sacrificing the spatial accuracy of IMRI data already at the network construction stage. The method yields results that summarize differences in connectivity at the module level, using a set of modules as a frame of reference across groups or conditions. In contrast to some alternative approaches for studying differences in module-level connectivity, the results produced by our method are easy to interpret and verify statistically – they allow to ‘see the forest for the trees’. Because data on the structural and functional human connectome are becoming more and more detailed, we believe such methods will play an increasingly important role in understanding the module-level structure of functional networks.

**Supporting Information**

Additional supporting information can be found in the online version of this article:
- Data S1 Labeling of the consensus modules
- Data S2 Node labels
- Data S3 Number of modules and representativeness of the consensus partitions
- Data S4 Distributions of supra-threshold link distances
- Data S5 Average coarse-grained movie networks

**Acknowledgments**

Computing resources provided by the Aalto Science IT project are acknowledged. EG acknowledges aivoAALTO Project Grant from the Aalto University, doctoral program “Brain & Mind”. Support by the Academy of Finland is greatly acknowledged. We thank Onerva Korhonen for comments on the manuscript.

**Conflicts of interests**

The authors declare no conflicts of interest.
Hubert, L. & Arabie, P. (1985) Comparing partitions. *J. Classifi.*, 2, 193–218.

Karypis, G. & Kumar, V. (1998) A fast and high quality multilevel scheme for partitioning irregular graphs. *SIAM J. Comput.*, 20, 359–392.

Lahnakoski, J.M., Glerean, E., Salmi, J., Jääskeläinen, I.P., Sams, M., Hari, R. & Nummenmaa, L. (2012a) Naturalistic fMRI mapping reveals superior temporal sulcus as the hub for the distributed brain network for social perception. *Front. Hum. Neurosci.*, 6, 233.

Lahnakoski, J.M., Salmi, J., Jääskeläinen, I.P., Lampinen, J., Glerean, E., Tikka, P. & Sams, M. (2012b) Stimulus-related independent component and voxel-wise analysis of human brain activity during free viewing of a feature film. *PLoS One*, 7, e35215.

Lancichinetti, A. & Fortunato, S. (2012) Consensus clustering in complex networks. *Sci. Rep.*, 2, 336.

Lancichinetti, A., Radicchi, F. & Ramasco, J.J. (2010) Statistical significance of communities in networks. *Phys. Rev. E*, 81, 046110.

Lancichinetti, A., Radicchi, F., Ramasco, J.J. & Fortunato, S. (2011) Finding statistically significant communities in networks. *PLoS One*, 6, e18961.

Lord, A., Horn, D., Breakspear, M. & Walter, M. (2012) Changes in community structure of resting state functional connectivity in unipolar depression. *PLoS One*, 7, e41252.

Meilä, M. (2007) Comparing clusterings – an information based distance. *J. Multivariate Anal.*, 98, 873–895.

Mennes, M., Kelly, C., Colcombe, S., Castellanos, F.X. & Milham, M.P. (2013) The extrinsic and intrinsic functional architectures of the human brain are not equivalent. *Cereb. Cortex*, 23, 223–229.

Meskaldji, D.E., Ottet, M.-C., Cammoun, L., Hagmann, P., Meuli, R., Eliez, S., Thiran, J.P. & Morgenhalter, S. (2011) Adaptive strategy for the statistical analysis of connectomes. *PLoS One*, 6, e23009.

Meunier, D., Achard, S., Morcom, A. & Bullmore, E.T. (2009a) Age-related changes in modular organization of human brain functional networks. *Neuroimage*, 44, 715–729.

Meunier, D., Lambiotte, R., Fornito, A., Ersche, K.D. & Bullmore, E.T. (2009b) Hierarchical modularity in human brain functional networks. *Front. Neuroinform.*, 3, 37.

Meunier, D., Lambiotte, R. & Bullmore, E.T. (2010) Modular and hierarchically modular organization of brain networks. *Front. Neurosci.*, 4, 200.

Mousse, M.N., Steen, M.R., Laurienti, P.J. & Hayasaka, S. (2012) Consistency of network modules in resting-state fMRI connectome data. *PLoS One*, 7, e44428.

Newman, M. (2010) Networks: An Introduction. Oxford University Press, UK.

Newman, M.E.J. & Girvan, M. (2004) Finding and evaluating community structure in networks. *Phys. Rev. E*, 69, 26113.

Papo, D., Zanin, M., Pineda-Pardo, J.A., Boccaletti, S. & Buldú, J.M. (2014) Functional brain networks: great expectations, hard times and the big leap forward. *Philos. T. Roy. Soc. B*, 369, 20130525.

Power, J.D., Cohen, A.L., Dosen, M.W., Wig, G.S., Barnes, K.A., Church, J.A., Vogel, A.C., Laumann, T.O. et al. (2011) Functional network organization of the human brain. *Neuron*, 72, 665–678.

Power, J.D., Barnes, K.A., Snyder, A.Z., Schlaggar, B.L. & Petersen, S.E. (2012) Spurious but systematic correlations in functional connectivity mri networks arise from subject motion. *Neuroimage*, 59, 2142–2154.

Power, J.D., Mitra, A., Laumann, T.O., Snyder, A.Z., Schlaggar, B.L. & Petersen, S.E. (2014) Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage*, 84, 320–341.

Reichardt, J. & Bornholdt, S. (2006) Statistical mechanics of community detection. *Phys. Rev. E*, 74, 016110.

Rosvall, M. & Bergstrom, C.T. (2010) Mapping change in large networks. *PLoS One*, 5, e8694.

Salmi, J., Glerean, E., Jääskeläinen, I.P., Lahnakoski, J.M., Kettunen, J., Lampinen, J., Tikka, P. & Sams, M. (2014) Posterior parietal cortex activity reflects the significance of others’ actions during natural viewing. *Hum. Brain Mapp.*, 35, 1762–1777.

Salvador, S., Suckling, J., Coleman, M.R., Pickard, J.D., Menon, D. & Bullmore, E. (2005) Neurophysiological architecture of functional magnetic resonance images of human brain. *Cereb. Cortex*, 15, 1332–1342.

Smith, S.M., Fox, P.T., Miller, K.L., Glahn, D.C., Fox, P.M., Mackay, C.E., Filippini, N., Watkins, K.E. et al. (2009) Correspondence of the brain’s functional architecture during activation and rest. *Proc. Natl. Acad. Sci. USA*, 106, 13040–13045.
Smith, S.M., Miller, K.L., Salimi-Khorshidi, G., Webster, M., Beckmann, C.F., Nichols, T.E., Ramsey, J.D. & Woolrich, M.W. (2011) Network modelling methods for fMRI. *NeuroImage*, 54, 875–891.

Sporns, O. & Betzel, R.F. (2016) Modular brain networks. *Annu. Rev. Psychol.*, 67, 613–640.

Stanley, M.L., Moussa, M.N., Paolini, B.M., Lyday, R.G., Burdette, J.H. & Laurienti, P.J. (2013) Defining nodes in complex brain networks. *Front. Comput. Neurosci.*, 7, 169.

Steen, M., Hayasaka, S., Joyce, K. & Laurienti, P. (2011) Assessing the consistency of community structure in complex networks. *Phys. Rev. E*, 84, 016111.

Strehl, A. & Ghosh, J. (2003) Cluster ensembles — a knowledge reuse framework for combining multiple partitions. *J. Mach. Learn. Res.*, 3, 583–617.

Uehara, T., Tobimatsu, S., Kan, S. & Miyachi, S. (2012) Modular organization of intrinsic brain networks: A graph theoretical analysis of resting-state fMRI. In Proceedings of the International Conference on Complex Medical Engineering 2012, pages 722–727. IEEE.

Watts, D.J. & Strogatz, S.H. (1998) Collective dynamics of ‘small-world’ networks. *Nature*, 393, 440–442.

Yacoub, E., Harel, N. & Ugurbil, K. (2008) High-field fmri unveils orientation columns in humans. *Proc. Natl. Acad. Sci. USA*, 105, 10607–10612.

Yeo, B.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W. et al. (2011) The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.*, 106, 1125–1165.

Zalesky, A., Fornito, A. & Bullmore, E.T. (2010) Network-based statistic: identifying differences in brain networks. *NeuroImage*, 53, 1197–1207.

Zalesky, A., Cocchi, L., Fornito, A., Murray, M.M. & Bullmore, E. (2012) Connectivity differences in brain networks. *NeuroImage*, 60, 1055–1062.