Community detection in weighted brain connectivity networks beyond the resolution limit

Carlo Nicolini[a,b], Cécile Bordier[a], Angelo Bifone[a,*]

[a] Center for Neuroscience and Cognitive Systems, Istituto Italiano di Tecnologia, Rovereto (TN), Italy
[b] University of Verona, Verona, Italy

ABSTRACT

Graph theory provides a powerful framework to investigate brain functional connectivity networks and their modular organization. However, most graph-based methods suffer from a fundamental resolution limit that may have affected previous studies and prevented detection of modules, or "communities", that are smaller than a specific scale. Surprise, a resolution-limit-free function rooted in discrete probability theory, has been recently introduced and applied to brain networks, revealing a wide size-distribution of functional modules (Nicolini and Bifone, 2016), in contrast with many previous reports. However, the use of Surprise is limited to binary networks, while brain networks are intrinsically weighted, reflecting a continuous distribution of connectivity strengths between different brain regions. Here, we propose Asymptotical Surprise, a continuous version of Surprise, for the study of weighted brain connectivity networks, and validate this approach in synthetic networks endowed with a ground-truth modular structure. We compare Asymptotical Surprise with leading community detection methods currently in use and show its superior sensitivity in the detection of small modules even in the presence of noise and intersubject variability such as those observed in fMRI data. We apply our novel approach to functional connectivity networks from resting state fMRI experiments, and demonstrate a heterogeneous modular organization, with a wide distribution of clusters spanning multiple scales. Finally, we discuss the implications of these findings for the identification of connector hubs, the brain regions responsible for the integration of the different network elements, showing that the improved resolution afforded by Asymptotical Surprise leads to a different classification compared to current methods.

1. Introduction

The brain is thought to consist of a network of interconnected, interacting components whose architecture is critical for the emergence of adaptive behaviors and cognition (McIntosh, 2000). Graph theory provides a powerful means to assess topology and organization of brain connectivity networks, like those derived from MRI and other neuroimaging methods (Eguíluz et al., 2005; Bullmore and Sporns, 2009). Within this framework, the brain is represented as a network of n nodes interconnected by m links. Typically, the nodes correspond to anatomically defined brain regions and the links to a measure of interregional interaction or similarity (Bullmore and Sporns, 2009). For resting state functional connectivity networks, edge weights are defined as interregional temporal correlations in the fluctuations of the BOLD signals, and the resulting graph can be represented by a correlation adjacency matrix. The arcs of structural connectivity networks (the "connectome"), conversely, reflect the number of white matter tracts connecting any two regions. Brain networks have also

been defined on the basis of intersubject anatomical covariance (Evans, 2013), co-activation of different brain regions across individuals subjected to experimental tasks (Crossley et al., 2013) or pharmacological challenges (Schwarz et al., 2007, 2008). All of these networks are "weighted" by definition, i.e. their edges are associated with real numbers representing a measure of the strength of pairwise interactions between nodes.

Graph-theoretical analysis of these networks has contributed substantially to our understanding of the topological organization of brain connectivity, revealing a small-world, rich-club structure (Eguíluz et al., 2005; van den Heuvel and Sporns, 2011) and the presence of hub regions characterized by high connectivity and network centrality. Additionally, a number of studies (reviewed in Bullmore and Sporns (2009) and van den Heuvel and Hulshoff Pol (2010)) have investigated the modular structure of brain connectivity networks, highlighting cohesive clusters of nodes that are more densely connected among themselves than with the rest of the network. In the graph-theory jargon, these disjoint clusters are sometimes dubbed "communities",

remnant of early investigations in the field of social sciences (Girvan and Newman, 2002).

Topological modularity is thought to reflect functional and anatomical segregation, a feature that may confer robustness and adaptivity to brain networks. Moreover, the degree of clustering within functional connectivity graphs may provide a measure of the balance between segregation and integration underlying brain function (Bullmore and Sporns, 2009). Finally, the identification of modules and their boundaries is important to understand the topological function of hub regions within the network (Meunier et al., 2010). Indeed, hubs sharing a large number of within-module edges may be critical to determine segregation of sub-structures within the network, while hubs connecting different modules are responsible for network integration (van den Heuvel and Sporns, 2013). It has been suggested that hub regions may have increased susceptibility to the effects of brain disorders (Crossley et al., 2014; Buckner et al., 2009), and their correct identification and classification are important to understand their putative role in the spread and manifestation of brain disease. Finally, alterations in the community structure of the brain have been observed in several neuropsychiatric conditions, including Alzheimer disease (Tijms et al., 2013), schizophrenia (Stam, 2014) and chronic pain (Balenzuela et al., 2010), and assessment of the brain modular organization may provide a key to understanding the relation between aberrant connectivity and brain disease.

Following initial work by Hilgetag et al. (2000), several graph theoretical methods have been deployed to investigate the modular structure of brain networks (Meunier et al., 2010, 2010; Power et al., 2011). Typically, these methods rely on the optimization of a fitness function that measures the quality of a network partition against that of an ensemble of randomized networks with similar statistical properties (the "null model"). Optimization of the fitness function of choice is often computationally demanding and scales steeply with increasing network size. Hence, heuristics are needed to calculate nearly optimal partitions of large networks, like those derived from neuroimaging data, within reasonable computation time (Blondel et al., 2008; Rosvall and Bergstrom, 2008).

A seminal finding in graph theory is that clustering methods based on optimization of a global function suffer from a resolution limit (Fortunato and Barthélemy, 2007), as they are unable to resolve modules that are smaller than a scale determined by the size of the entire network. This problem was first demonstrated for Newman's Modularity (Newman, 2006), a method included in the Brain Connectivity Toolbox (Rubinov and Sporns, 2010) and most frequently applied to the analysis of neuroimaging data.

Subsequent work by various groups has demonstrated that the resolution limit is quite pervasive and affects, to a different extent, many other methods based on optimization of a global fitness functions (Squartini et al., 2015; Traag et al., 2011; Lancichinetti and Fortunato, 2009), including Reichardt and Bornholdt's (2006), Arenas et al. (2008), Ronhovde and Nussinov (2009), Rosvall and Bergstrom's (Infomap) (Rosvall and Bergstrom, 2008; Kawamoto and Rosvall, 2015) and others.

The introduction of a resolution parameter has been proposed as a means to mitigate the problem by adjusting the resolving power of the function to a specific scale (Reichardt and Bornholdt, 2006; Ronhovde and Nussinov, 2010; Thomas Yeo et al., 2011). However, this approach enables resolution of smaller clusters at the expense of larger ones, which may be unduly subdivided, thus resulting in partitions with relatively uniform cluster size distributions that do not capture the complex modular structure of real-world networks (Lancichinetti and Fortunato, 2011).

Recently, we have assessed the effects of the resolution limit on the analysis of brain connectivity networks (Nicolini and Bifone, 2016). Specifically, we have shown that this limitation severely curtails the ability to detect small, but functionally and anatomically meaningful clusters of nodes even when they present high densities of intra-cluster

edges. Moreover, we showed that resolution-limited methods, like Newman's Modularity, do not reflect the multiple scales of the organization of brain connectivity networks, where small and large modules can coexist. We have also demonstrated that Surprise, a conceptually different fitness function grounded in probability theory, behaves like a resolution-limit-free function (Nicolini and Bifone, 2016). Maximization of Surprise, based on an algorithm dubbed FAGSO, revealed a heterogeneous distribution of modules within brain resting state and coactivation networks. If confirmed, these findings would suggest that a substantial revision of current models of brain modular structure may be in order.

A fundamental limitation of Surprise lies in its definition in terms of discrete probability and binomial coefficients that make it applicable only to binary networks, i.e. graphs with edge values 1 or 0. This may represent a substantial drawback, for it requires binarization of brain connectivity networks, thus discarding potentially important information contained in the edge weight distribution. Moreover, different binarization procedures may lead to different network representations for the same connectivity dataset. Therefore, an extension of Surprise to weighted networks would be highly desirable, and would provide a new and important tool to study the modular organization of brain connectivity beyond the resolution limit.

Capitalizing on recent development in the field of statistical physics of complex networks (Traag et al., 2015), here we describe and demonstrate the use of Asymptotical Surprise, a weighted counterpart to Surprise, in the study of the modular structure of weighted networks. Moreover, we propose a new algorithm, dubbed PACO (PArtitioning Cost Optimization) for the maximization of Asymptotical Surprise.

Since there is no ground-truth structure for brain functional connectivity networks, we have assessed the performance of this novel approach on synthetic networks with a planted modular structures, and compared it to some of the leading graph partitioning methods. Importantly, we demonstrate our approach in networks derived from synthetic data that mimic different structures, levels of noise and variability, such as those observed in functional connectivity experimental data. Indeed, improved resolution afforded by Asymptotical Surprise may imply increased vulnerability to spurious modules resulting from noisy correlations. It is therefore important to assess the benefits of increased resolution against the limitations arising from intrinsic data variability.

Finally, we apply Asymptotical Surprise to weighted functional connectivity networks from resting state fMRI data, revealing a heterogeneous, multiscale community structure. We show that the finer modular subdivision of resting state functional connectivity networks obtained by Asymptotical Surprise leads to substantial differences in the identification of connector hubs compared to other community detection methods.

2. Materials and methods

2.1. Notation

Here we briefly summarize the terminology and the notation that will be used throughout the paper. A binary graph $G = (V, E)$ is a representation of a set V of n nodes, also called vertices, connected by m links (or edges), in a set E. The adjacency matrix $A = \{a_{ij}\}$ of a binary graph is a square $n \times n$ symmetric matrix with elements $A_{ij} = 1$ when an edge exists between vertex i and j and 0 otherwise. We denote the total number of possible links in the graph as $p = \binom{n}{2}$.

A weighted graph $G = (V, E, W)$ assigns as a set of edge weights W to the links. For weighted graphs, the adjacency matrix is square, symmetrical and has real elements.

A clustering $\zeta = \{\zeta_c\}$ of G is a partitioning of V into disjoint sets of nodes, $\zeta_c \subseteq V$, which we call modules or communities. Each module consists of n_c nodes, m_c edges and $p_c = \binom{n_c}{2}$ pairs of nodes. On