A network neuroscience of neurofeedback for clinical translation

Andrew C. Murphy[a,b] and Danielle S. Bassett[a,c]

Abstract

In the emerging field of network neuroscience, the brain is represented as a network of discrete yet functionally and structurally interconnected areas. Mathematical and computational tools to characterize the organization of this network can provide insights into the principles subserving brain structure and function, and can pinpoint differences between healthy individuals and individuals suffering from psychiatric disease or neurological disorders. The field is now faced with the question of how to devise clinical interventions that target these network alterations. Potential solutions to this question include the combination of emerging theories of network control with cutting-edge interventions such as neurofeedback. Each of these techniques may now be mature enough to combine to obtain a theoretically-motivated framework informing viable neuropsychiatric therapies.

Addresses

[a] Department of Bioengineering, University of Pennsylvania, Philadelphia, PA 19104, USA
[b] Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
[c] Department of Electrical & Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104, USA

Corresponding author: Bassett, Danielle S (dsb@seas.upenn.edu)

Introduction

The human brain is fundamentally a network, or a system of interconnected functional units [1]. Such a network can be formulated as a graph, a mathematically well-defined object that is amenable to empirical study. A brain graph represents areas or regions as network nodes, and it represents connections between those areas as network edges (Fig. 1) [2]. The graph representation of the human brain is conceptually flexible: inter-regional connections can either be structural in nature — for example, estimating the number or volume of white matter tracts between areas — or they can be functional in nature — for example, estimating the degree of functional influence that one area has on another. In the context of both structure and function, network representations have proven particularly useful in deriving organizational principles at a systems level [3].

Importantly, the network approach can also be used to better understand when these organizational principles are altered, or when the normative generative mechanisms supporting those principles go awry [4]. Indeed, by facilitating a comparison of healthy and diseased brains, the network approach has led to the discovery of characteristic differences underlying many neuropsychiatric diseases and neurological disorders [5], which may be responsible for the observed clinical symptoms. Despite these successes, it is not yet common practice to exploit network differences to devise viable clinical therapies. Establishing such a framework would offer a bottom-up approach, informing the development of neuromodulation therapies directly with features of the interconnected brain itself, to address the network-level functional alterations that characterize diseased brains.

The development of such a framework is challenging on many levels. For example, while it is often clear how a diseased brain network functionally differs from a healthy brain network, it is not well understood what functional alterations must be induced in the diseased network in order to reconfigure it into the healthy state [6]. To address this challenge, we must have a mechanistic, dynamic understanding of how the network functions. With such an understanding, we may be able to determine how a given functional alteration to the network impacts its function in the future, for example by bringing a diseased brain network back in line with how it is configured in healthy brains. This dynamic, mechanistic view of networks, known as network control theory (Fig. 2), is an established framework in other areas of network science and engineering [7], and is primed for extension into neuroengineering more generally and network neuroscience specifically [8]. Technical challenges must be addressed before clinically applying control theory, but those are not discussed here.

Moreover, once the necessary network alteration is understood, there must exist a means with which to induce such a change. In the context of neuropsychiatric

Figure 1
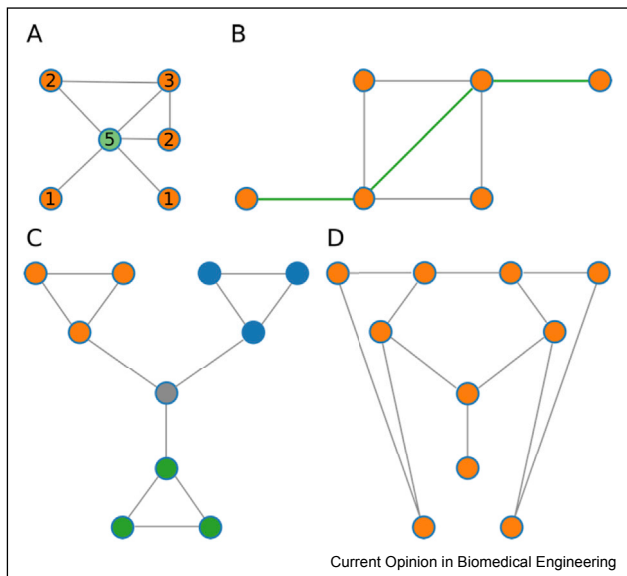
Statistics to quantitatively characterize network organization in the human brain. (A) A schematic of node degree. The degree of each node is indicated by a number superimposed on top of it; the strongest hub node is located in the center of the graph, with degree 5, highlighted in green. (B) The shortest path length between two nodes is the path that traverses the fewest intervening nodes. In this schematic, the shortest path between the two outer nodes is highlighted in green. (C) An example of a modular network in which there are three groups of densely interconnected nodes. One module contains the orange nodes, another the blue nodes, and a third the green nodes; the grey node is not part of a module. (D) A non-modular network, to offer a contrast to the graph in panel (C).

disorders, this network alteration may be altered (higher or lower) functional connectivity between particular brain regions, the specifics of which may be disorder-dependent. To address this challenge, we consider the viability of neurofeedback as a mechanism to provide therapeutic brain network alterations [9]. Neurofeedback is a type of biofeedback wherein the subject is presented, in real time, with information derived from brain signals. The signals, which originate from electroencephalographic (EEG), functional magnetic resonance imaging (fMRI), or others imaging modalities, are presented directly to the subject, typically in audio or video format. This forms a closed-loop system that enables the subject to regulate their own brain activity (Fig. 3). Using neurofeedback, a subject may be taught to stimulate a specific area or connection in their brain [10] that is thought, based on a mechanistic dynamic model, to be likely to create alterations in their brain network that may attenuate symptomatology.

In this brief article, we discuss the current state of the field supporting the development of neuromodulation therapies informed by brain network structure and function. We begin with a brief account of the application of tools from network science to neuroimaging data to better understand the brain from a systems perspective. We recount seminal studies revealing network-level alterations in brain connectivity in disease states, and then we describe engineering-based theories of driving a network from an altered state to a healthy state via targeted interventions. We focus on findings from the past decade, with an emphasis on results from the past 2 years. We consider how this control framework might lead to the principled development of therapeutic interventions that capitalize on neurofeedback. Finally, we suggest that such a mechanistically motivated, bottom-up approach to neurotherapy has significant clinical advantages over the current manner in which therapeutic neuromodulation is delivered.

Network neuroscience

Essential to any therapeutic methodology is the correct identification of the target on which to intervene. In clinical scenarios when the target is a pattern of distributed changes in the structure or function of neural circuits, the tools of network neuroscience may be particularly useful. The unique contribution of the network approach lies in its ability to identify potentially complex patterns of functional differences in the brains of a healthy *versus* target population, rather than narrowly searching for univariate discriminative features. Capitalizing on this sensitivity to distributed features could enable the development of a therapy to correct those differences.

Network neuroscience is a mathematical, computational, and theoretical framework with which to understand the structure and function of the brain, and the interactions between its component parts [11,12]. At the most basic level, network neuroscience views the brain as a system of nodes connected by edges. Traditionally, nodes are areas or regions of interest, often selected by parcelling the brain into a set of discrete non-overlapping volumes. Edges connecting nodes can represent structural relationships [2], for example defined by the number or strength of white matter tracts linking two brain areas [13]. Alternatively, edges can represent functional relationships, for example defined by the magnitude of temporal correlation between the activity of two brain areas [14]. The network approach breaks from the classical univariate view of brain activity by placing emphasis on the patterns of interactions *between* brain areas, rather than on regional activity alone. In this way, we study functional connectivity patterns in isolation, without considering regional activation, structural connectivity, or other types of information.

Interaction patterns can be quantitatively characterized using topological statistics to summarize networks in