Step-wise target controllability of driver nodes in biological networks

Giulia Bassignana

Sorbonne Université

Ecole Doctorale Informatique, Télécommunications et Electronique
ARAMIS LAB at the Paris Brain Institute

Step-wise target controllability of driver nodes in biological networks

Giulia Bassignana

Doctoral thesis in Informatics

Supervised by Fabrizio De Vico Fallani
and co-supervised by Violetta Zujovic *and* Olivier Colliot

To be defended on 30/10/2020

Members of the jury:
- Colliot Olivier, HDR, thesis co-supervisor
- Commault Christian, full professor, HDR, rapporteur
- De Vico Fallani Fabrizio, HDR, thesis supervisor
- Deveaux Frederic, full professor, HDR, examinateur
- Frasca Mattia, associate professor, rapporteur
- Zujovic Violetta, HDR, thesis co-supervisor

Acknowledgements

I would like to start by thanking my advisors Violetta Zujovic, Olivier Colliot, and Fabrizio De Vico Fallani for their teaching, patience, and guidance over the years. In particular, I would like to thank Fabrizio for his ability to see the best of every idea, and for always being positive and supportive.

I am grateful to the members of the jury Christian Commault, Mattia Frasca, and Frédéric Devaux for the advice and attention in reading and providing feedback for the thesis, helping improving it.

It has been a pleasure to interact and work with all the members of the ARAMIS team, who have been of great assistance in getting feedback on ideas and presentations, taking the mind off things, and starting to learn Spanish. I would like to thank all the people inside and outside the ICM with whom I shared a part of my path over the past three years.

A special thanks to Tiziana, who is like a twin sister to me, to Iacopo and Raffaele, who were always ready to share PhD experiences, and to all my friends related to 'Casa PCS', who are an amazing and never-ending source of physics talks and good food.

Finally I would like to thank my family who have been nothing but supportive over the years, and Stefano, who I know is always there for me.

Abstract

The possibility of using mathematical tools to describe and influence complex interconnected systems is getting more and more attainable. Methods based on network controllability to identify the nodes able to impact the state of a whole system are nowadays increasingly studied. However, the problem has a high combinatorial and numerical complexity because of the huge number of a priori equivalent solutions. There has recently been a growing interest in finding the minimum number of inputs to control the whole or a part of the system, and in evaluating the ability of a single node in steering this process. However, specific problems have drawn less attention. In some biological settings it may be required to act on a single node, and it may be of interest to affect only a well-defined subset of the units, a target set. This leads to a single input target control problem, where we can exploit biological constraints to study the relative importance of different driver nodes.

This dissertation aims to apply controllability theory to biological networks in an original way, to understand what insight mathematical controllability theory can bring to biological networks, and to study the importance of different driver nodes in controlling a target set. We develop a heuristic that we call step-wise target controllability, which measures the centrality of a driver node as the number of targets it can control and provides a controllable configuration of targets. We show that this method is efficient for sparse directed networks. To test the theory in two empirical settings, we apply it to macrophage activation and to the aging of the brain. We find that this framework can identify dysregulated pathways in a molecular network of macrophage activation, and that it can characterize normal aging in the brain. Our method represents a practical answer to use to our advantage the complexity of the control problem, exploiting existing biological knowledge.

Keywords Controllability theory, network controllability, target controllability, driver nodes, biological networks, molecular networks, brain networks.

Contents

List of Figures

List of Tables

Acronyms

τ step-wise target control centrality. vii–x, xii, 23, 24, 34–36, 41, 45, 47, 48, 50–58, 62, 63, 81, 87–91

DMN default mode network. ix, 54, 57, 58

DTI diffusion tensor imaging. x, 49, 50, 58, 60, 83

fMRI functional magnetic resonance imaging. 2, 49, 50, 58, 60, 61

HC healthy control. vi–ix, xi, 30–33, 35–39, 42, 46, 47, 67, 71, 72, 75, 76

LTI linear time invariant. iii, 2, 6, 7, 15–19, 50, 59, 61, 64

MDS minimum driver node set. 9

MS multiple sclerosis. vi–ix, xi, 3, 30–33, 35–39, 42, 46, 47, 63, 67, 73, 74, 77, 78

PBH Popov-Belevitch-Hautus. 8, 16, 18, 25, 40, 64

ROI region of interest. ix, x, xii, 50, 52–55, 62, 80, 85, 86, 88–91

Chapter 1

Introduction

For thousands of years, humans have modified their environment to obtain better conditions. To change unwanted unhealthy outcomes into desired healthy states may be seen as the ultimate goal of human manipulation. In this regard, the possibility to apply mathematical controllability theory to influence complex systems is extremely attractive.

This dissertation aims to bring insights by applying mathematical controllability theory to biological networks in an original way. In particular, we intend to study the importance of different driver nodes in controlling a context-specific target set, and to test the theory in empirical settings. We study two applications: the activation of macrophage cells, and the aging of the brain.

Recently, complex networks have received more and more attention and have been studied in disparate fields such as finance [10, 46, 136], social sciences [73, 91, 133], ecology [49, 132, 180], machine learning [59, 64, 169], epidemiology [30, 60, 128]. This attention led to multidisciplinary efforts that made possible a rich exchange and cross-contamination of techniques from different fields. One example is the application of controllability, from linear systems theory, to complex networks [87, 102]. Roughly at the same time, in biology there has been an effort to go from reductionist approaches to integration and systems biology. This context promoted the representation of complex biological systems as networks [129]. To apply controllability theory to biological networks was then a natural step forward [65, 103].

In control theory, it is said that a system is controllable if it can be steered from any initial state to any final state in finite time, by applying an external input to well-chosen nodes, that are called drivers [82, 108, 153]. When the external input is applied to only one driver node, the controllability problem is called single-input. If the interest is to change the state of a subset of nodes, instead of that of the whole system, we call this subset target set.

Currently, biological network controllability focuses on two main directions. The first one is to find the minimum inputs required to control the whole or part of the system [103]. An issue with this approach is that in general many different sets of driver nodes can control the system [170]. The second direction is to evaluate the power of a single node in controlling the whole system [65, 104]. In this case, it is relatively easy to compute how many nodes can be controlled by a single driver, but it is not trivial to identify which subsets of nodes can be controlled simultaneously. Moreover, the problem of evaluating a single driver in controlling a target set has surprisingly received little attention. It will be one of the main points addressed in this thesis.

From an empirical and biological perspective, it is often of interest to manipulate the state of only a subset of the nodes in the network [187], and it may be extremely difficult to act on more than one node at a time, as in the case of gene knockout, and brain stimulation. Our main research question is to understand what insight can controllability theory bring to biological networks, and in particular what is the role of a single node in influencing a context-specific target set of the network. We tested our framework in two different biological applications: the activation of macrophages, and the aging of the brain.

We integrated network theory, linear systems theory, and biological knowledge to study the role of single driver nodes in controlling a context-specific target set, devising a measurable way to distinguish between a priori equivalent solutions. The main contribution of this dissertation is the development of a heuristic that we call step-wise target controllability [12]. This heuristic is shown to be efficient for sparse directed networks. Our framework is original in the fact that it is not centered on the identification of drivers and targets, but instead on finding hidden functional relations between drivers and their controllable targets.

We used methods from linear system theory, thus we required a linear time invariant setting, in which the dynamics of the system is linear and the structure of the underlying network does not change in time. We developed a heuristic step-wise algorithm that addresses the single input target control problem through an exact method that takes into account biological knowledge of the specific setting to lower the combinatorial complexity of the problem. We measured the importance of a driver node as the number of targets it can control. Additionally, we exploited context-specific information to assign an ordering to the target set, which made possible the identification of the highest-ranking controllable configuration of targets. In the case of macrophage activation, we studied function through gene expression, and in the case of the brain, we adopted functional magnetic resonance imaging

measurements.

We showed that this method is able to uncover hidden functional relations between drivers and their controllable targets. In the case of macrophage cells activation, we can identify dysregulated pathways and the genes that may be responsible for it. In the brain, we can characterize normal aging.

The dissertation proceeds as follows. **Chapter 2** presents an overview of network controllability, with a focus on biological applications. **Chapter 3** explains fundamental notions of graph theory and linear systems controllability theory. **Chapter 4** describes the step-wise target control centrality framework. **Chapter 5** presents the application of our method to macrophage cells activation in multiple sclerosis and to aging in brain networks. **Chapter 6** concludes the dissertation with a summary of the findings and addresses possible directions for future research.

Chapter 2

Network controllability: from engineering to biological systems

Network science is well-versed in multidisciplinary studies because of its ability to model the interaction of distinct units at different scales and levels of detail. Neural networks, transportation networks, electric power grids, the Internet, the World Wide Web, scientific coautorship and citations, genetic, metabolic, and protein networks are only some of the possible applications. Recently, there has been a growing effort in conducting multi-disciplinary projects, in the hope to achieve better understanding from different perspectives. This dissertation integrates multiple disciplines and in the following there will be a description of the fundamental concepts used. In particular, it is described how mathematical controllability theory - motivated by biological questions - was applied to networks in general, and to biological networks.

Network science started as a branch of discrete mathematics, called graph theory. We can say that it was born in 1736 when Euler solved the Königsberg's bridges problem. Since then there have been great advances in the study of network topology, and in the study of more and more complex systems. In general, we call a network complex when it presents some of the following qualities: it is large-scale, it has a higher-order dynamics, and it has an irregular connection pattern. In 1960, Erdös and Rényi [54] proposed the concept of random graphs. More recently, Watts and Strogatz [179] studied the topology of not purely random neither purely regular networks, developing the concept of small world graphs. One year later, Barabási and Albert [7] proposed the scale-free model, able to explain the important role played by high-degree nodes, the hubs.

The attention of the physics community and improved computational tools made possible the study of structural and topological network properties. Universal and unifying principles and statistical properties of graphs were identified [9]. The debate then shifted to growth models, because the evolution of a network and its change in time (caused by external or internal forces) impose constraints and limitations to function and behavior [18]. Because of the strong link between topology and function, the attention shifted to the interactions between fundamental units of the network, asking if there could exist an interplay between network topology and dynamics, and if there could be common patterns in the dynamics [11]. To answer these questions, perturbation studies and experiments were carried out. It was shown that perturbations of one node can alter the states of other nodes, due to the interaction between units [53, 85]. This property has been exploited to control a network.

Control theory is a branch of mathematics that studies the behavior of dynamical systems. The objective is to influence the state of the system by means of a feedback process: appropriately chosen inputs that are supposed to drive the the system to a target trajectory or final state. To bring the system's output to the desired one, a control action made of the difference between the actual and the desired output is applied as feedback to the input of the system [156]. The property that determines if a system is controllable is called *controllability*. It informs on whether and on which conditions control may be achieved. We say that a system is controllable if it can be driven from any initial state to any final state in finite time, by applying inputs to well-chosen nodes in the system, called *driver nodes* [82, 108]. Since controllability is a prerequisite for control, it is crucial to understand what factors determine it, in particular from a topological and structural perspective.

Mathematical control theory is widely applied in engineering systems. For instance in electric circuits, manufacturing processes, communication systems, aircraft and spacecraft, and coordination of moving robots. These physical systems can be mathematically described by means of a set of inputs, outputs, and state variables. These elements are linked by a set of differential equations in the state space representation [156]. Despite the fact that these equations are often nonlinear, control theory heavily relies on linear systems theory. This is mainly due to two factors. First, controllability of locally linearized systems is studied before the full nonlinear problem, which is particularly accurate for systems near their equilibrium points. Second, the complexity of the problem is increased by non-trivial network topologies, making it extremely difficult to handle a nonlinear setting. Instead, linear controllability can be seen as a prerequisite of nonlinear controllability. Moreover,

in some real problems such as consensus, the linear system represents an accurate descriptions and provides valid solutions [102].

Linearization of nonlinear systems around and equilibrium point leads to a linear time invariant (LTI) system, where the structure of the system does not change in time. LTI systems are a sub-problem of linear systems. This dynamics was used to model for instance the amount of traffic that passes through a node on a communication network, or transcription factor concentration in a gene regulatory network. A classical criterion to test for controllability in a LTI system for a given set of inputs is the Kalman rank condition, that relies on the controllability matrix of a system. It will be discussed in detail in **Chapter 3**.

Classical control problems typically take into account a single higher-dimensional system. In such problems the complexity may be given by high dimensionality, a stochastic setting, and a nonlinear dynamics. The difficulty of the problem makes it not yet feasible to study a directed and irregularly connected network of many such systems. In the last decades there has been a shift in the focus of control problems, mainly motivated by technical reasons that a technological advance has brought forward. Rapid development of network science and engineering highly improved available resources such as supercomputers, huge databases, GPS services, and cloud computational environments, needed by new necessities from systems such as the Internet, the World Wide Web, wireless communication networks, power grids, global transportation systems, genome-scale metabolic networks, protein interaction networks, and gene regulatory networks [26]. Currently, the focus is on the study of many dynamical systems in a networked framework, where great attention is paid to directionality and connectivity. The complexity in this case may be given by the large number of interacting units, the high dimensionality, the sophisticated and irregular connection pattern, the nonlinear dynamics, the time-varying environment, the presence of multiple spatio-temporal scales.

In this setting, it is evident that it would be unfeasible and too expensive to control all the nodes in the network to achieve a certain desired objective. Wang and Chen [178] and Li et al. [97] introduced pinning control, a new strategy that takes into account both dynamical rules and network topology in control systems design and implementation. The aim of this strategy is to control the whole dynamical network by directly imposing controllers to only a fraction of selected nodes in the network. Pinning control was inspired and motivated by biology, in fact in nature it was observed that a small fraction of individuals is able to influence the behavior of a whole group [26]. This is true for instance for foraging bees and fish schools [37].

This framework led to a new set of questions. In particular, it was asked how

many and which nodes in a network are needed to achieve the control goal most effectively. It is now clear that the answer to the above question depends on the structure of the network as well as on the dynamical laws governing it. Modern control theory studies the relation between control inputs and system state with a particular attention to the interplay between network topology and dynamics.

In general, we may say that to understand the behavior of a system we are required to measure its state variables and to model its dynamics [102, 103]. Ideally, we should perfectly know: i) the architecture of the system, that is its structure and interconnection pattern; ii) the dynamical laws governing it; and iii) all the system's parameters. In practice of course this is not the case, and we lack an accurate description of one or more of these three factors. For instance, it was estimated that in protein-protein interaction networks we are currently able to infer about 20% of all the plausible interactions [146]; an analytical form of the dynamics in communication systems is often missing; and accurate values of biochemical reaction systems are extremely difficult to retrieve [102]. It is also true that recently network controllability greatly advanced despite the lack of this ideal and complete knowledge. This may have also been favored by the rich cross-contamination from other fields. A brilliant example of this is the case of structural controllability.

Structural controllability of complex networks was first introduced by Lin [98]; it refers to the property of a system to be controllable for almost all of its numerical realizations [32], bringing attention to the topology. In particular, it evaluates the controllability of binary networks, in which edges are either present, or absent, effectively overcoming the need for estimating the system parameters. At the time when this concept was introduced, sufficient and necessary conditions to check whether a system with LTI dynamics was controllable were given. However, an algorithm to compute it was not provided.

Later, generic properties of structured systems were studied, allowing for a mapping of system's properties into graph theoretical conditions [33, 47, 172]. Liu et al. [103] integrated knowledge from statistical physics, spin glass theory and network science to map the structural controllability problem into a well-known graph theoretical problem: the maximum cardinality matching. They were able to provide answers to questions such as: what is the minimum set of nodes that should be controlled to achieve control over the system? How to estimate the fraction of such driver nodes?

Notice that, despite the claim that topology and degree distribution play a major role in network controllability [103], other authors stressed the importance of also taking into account the nodal dynamics [38, 176, 196]. Exact controllability crite-

ria such as the Kalman rank condition, the Popov-Belevitch-Hautus test, and the Gramian criterion - all discussed in **Chapter 3** - are limited by shortcomings such as parameter estimation and ill-conditioned numerical problems. However, they continue to be investigated because - unlike the structural controllability framework - they allow for the computation of the required control energy and other metrics [130].

In some settings, it may not be feasible to control the whole network. For instance, it was predicted that for gene regulatory networks over 80% of the nodes should be drivers in order to achieve control over the system [83, 103]. Moreover, in some applications it is not even necessary to control the whole network. This may be the case in smart grids, for the installation of controllers and for choosing the locations of actuators with physical limitations; and in the development of drug cocktails for multiple drug target identification in gene regulatory networks [184, 189].

To describe such problems, we can define a target control problem in which the aim is to find the minimum number of driver nodes required to steer the state of only a selected subset of the network's nodes, a subsystem. These nodes are called *target nodes*, and are chosen because they are supposedly essential for a certain task or mission of the system. The target control problem for linear networks is a particular case of output control, and a generalization of the full control problem, which requires the control over the entire system [40].

The multi-input target control problem was shown to be NP-hard [40], thus heuristics and approximate solutions have been proposed to solve it. Among the firsts to recently address this problem, Gao et al. [58] proposed a structural approach in the form of a greedy algorithm. Later, Czeizler et al. [40] built on their work improving the algorithm. A different perspective on the problem was proposed by Klickstein et al. [88], who focused on the control energy, and implemented a target problem with the aim to reduce the energy, without changing the predesignated driver nodes; they studied the relation between control energy and cardinality of the target set. Variations of the problem such as strong structural controllability [172], symmetric networks [95], and functional target controllability [35], where the interest is in controlling the whole trajectory instead of only the final state, were recently proposed. In all those cases, necessary and sufficient graph conditions for target controllability were provided.

It is important to notice that a trivial yet legit objection arises: would it be easier to directly control the target nodes? In practice, that may seldom be the case, in fact we may have a large and prohibitive number of targets, and the target nodes may not be directly accessible via a control action.

The main goal of the control problems discussed so far was to identify a minimum driver node set (MDS) able to achieve control over all or part of a system. It had been shown that the minimum number of drivers required for control, that is the cardinality of the MDS is fixed [33, 103], however the individual nodes in the set may vary [80, 81, 88]. It remains an open question to enumerate all possible minimum driver node sets, largely due to the huge number of maximum cardinality matchings [170], even in simple graphs such as trees [75, 174], bipartite graphs [101], or random graphs [192]. It is important to notice that since the MDS is not unique, nodes may not participate in control equally, prompting works that quantify their participation [31, 32, 80, 81]. This observation motivated the investigation of the composition of different MDSs, and the classification of the role of individual driver nodes in exercising control.

Commault et al. [34] worked on a classification for observability - the dual of controllability - and later extended their results to account for controllability with the hypothesis that each driver receives an independent input, also allowing for a case in which not all nodes can directly receive a control input [32]. Driver nodes were classified as essential, if they appeared in all configurations; useful, if they appeared in some configurations; and useless, if they were never present in a configuration.

Jia and Barabási [80] studied the case in which a single input signal can be applied to multiple nodes. They devised a classification based on the likelihood of each driver node to participate in a MDS. They estimated this probability with a random sampling algorithm, nodes that are always (never) present in the MDSs are called critical (redundant), and ordinary if they belong to the set sometimes.

Participation in a MDS could give insight into the role of a driver when sharing control with other nodes, but it does not answer to a more fundamental question of its role in controlling the system in the case in which it is the only driver node.

Typically in network theory the importance of a node is computed by means of centrality measures based on topology. Some examples are the degree, the betweennes, and the eigenvector centrality [124]. A measure of centrality based on controllability theory was proposed by Liu et al. [104]. Their *control centrality* is quantified as the size of the controllable subspace when a certain driver node is considered, and it can be computed through an optimization problem.

The single-input case may be particularly useful in biological contexts, where it could be unfeasible or extremely difficult to act on multiple driver nodes.

Gu et al. [65] studied structural brain networks, and for single brain areas proposed multiple measures of importance based on the Gramian matrix (defined in **Chapter 3**).

Recent developments in network theory, together with advances in computational tools and medical techniques made possible the study of biological systems from new, more detailed perspectives. Moreover, there has been a growing interest in applying control theory to biological systems. This was partially motivated by the intuitive fact that control and regulation are key-factors for biological systems [45, 77], which - for instance - are able to maintain homeostasis, sensing and processing external and internal cues in changing environmental conditions. In the evolution of molecular interaction networks, it may be particularly relevant the need to implement control in the form of input signals steering the network to specific functional states [173].

At a cellular level, control theory was applied to biological systems in the form of experimental and computational approaches that made possible the regulation of protein expression in a population of yeast cells [116], and to maintain a bistable genetic circuit to unstable equilibrium [109].

At large scale, such detailed control is not yet possible. However, recent studies were able to apply controllability theory to different kinds of biological networks, gaining substantial insight. In protein-protein interaction networks, Vinayagam et al. [173] exploited controllability theory to identify drug targets; Ravindran et al. [138] studied the virus-host interaction networks to understand the virus exploitation of the host system; Ravindran et al. [137] analyzed the cancer signaling network to propose therapeutic targets. Moreover, Yan et al. [187] performed experimental studies to test the predictions of controllability theory in the case of individual neurons and neuronal classes in the locomotor behavior of the *C. Elegans* nematode.

Studying the controllability of biological systems, a particular attention has been paid to the brain, a complex system that operates at different spatial and temporal scales. It enables humans to perform complex cognitive functions trough the alteration of the dynamics of neural systems in order to meet task demands. These capabilities are referred to as cognitive control [20, 72], and cognitive function is driven by dynamical interactions between large-scale neural circuits, enabling behavior.

At a system-level, advances in network theory and in the study of the topological properties of brain networks involve questions concerning the relation between structure and function, and how the evolution of the architecture of the brain is reflected in its functioning [57]. From a control-theoretical perspective, researchers focused on studying the controllability of the brain, and the brain areas that are most influential in driving changes to brain dynamics. Studies of non-invasive transcranial magnetic stimulation [16] and invasive deep brain stimulation [181] were

applied to healthy and diseased subjects, and are used for clinical treatments, as in the case of Parkinson's and Alzheimer diseases, epilepsy and stroke [119].

In conclusion, advances in network theory computational tools, and available resources, motivated by biological questions, prompted the study of biological systems from a networked, control-theoretical perspective.

Chapter 3

State-of-the-art on network controllability

We wish to understand what insight can mathematical controllability theory bring to biological networks, to study the importance of single driver nodes in controlling a context-specific target set, and to test the theory in empirical settings. To do so, we devised a measure of importance of the driver nodes, analyzing in an original way the control they may exert on target nodes in two biological applications: the activation of macrophage cells, and the aging brain.

In this chapter, we report fundamental notions of graph theory and controllability theory.

3.1 Graph theory

Graph theory is the natural framework for the mathematical treatment of complex networks [18]. Many objects of interest in the physical, biological, and social sciences can be modeled as networks and thinking of them in this way can lead to new and useful insights [57, 124].

A *graph* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consists of two sets $\mathcal{V}$ and $\mathcal{E}$, such that $\mathcal{V} \neq \emptyset$ and $\mathcal{E}$ is a set of either unordered for undirected graphs, or ordered for directed graphs, pairs of elements of $\mathcal{V}$. The elements of $\mathcal{V} \equiv \{v_1, ..., v_N\}$ are the *nodes*, or *vertices*, of the graph $\mathcal{G}$, while the elements of $\mathcal{E} \equiv \{l_1, ..., l_E\}$ are its *links*, or *edges*.

A *weighted* graph is a graph in which a number, the weight, is assigned to each edge. A *binary*, or *unweighted*, graph only accounts for the existence of the connection, and not for its strength.

In a *undirected*, or *symmetric*, graph, each of the edges is defined by a pair of

nodes i and j , and is denoted as (i, j) or l_{ij}. The edge is said to be *incident* in nodes i and j, or to *join* the two nodes; the two nodes i and j are referred to as the end-nodes of link (i, j).

In a *directed* graph, the order of the two nodes is important: l_{ij} stands for a link from j to i , and $l_{ij} \neq l_{ji}$. Edge l_{ij} is said to be *incident* in node i; node i is the end-nodes of link (i, j).

Two nodes joined by a link are referred to as *adjacent* or *neighboring*.

An *isolated* or *disconnected* node is a node that has no neighbors.

The *adjacency matrix* is a square matrix used to represent a finite graph. The elements of the matrix indicate whether pairs of nodes are connected or not in the graph. The elements of the adjacency matrix are either zeros and ones, for binary networks, or real numbers, for weighted networks. If the graph is undirected, the adjacency matrix is symmetric.

We call a *walk* a set of edges in an undirected graph that is ordered to form a sequence in which any pair of successive edges shares a common node. In a directed graph, the definition is analogous, but with the added restriction that the edges have to be directed in the same direction. A walk can traverse multiple edges and nodes, visiting the same node or edge on multiple occasions.

A *closed walk*, or *cycle*, is a walk in which the first and last nodes are the same.

A graph is said to be *acyclic* if it does not contain any cycle.

A *path* is a walk in which all edges and nodes are unique.

We call a *tree* an undirected graph in which any two nodes are connected by exactly one path [124]. A *rooted directed tree*, or *tree*, is a directed graph in which, for a vertex i called the root and any other vertex j, there is exactly one directed path from i to j.

A *subgraph* of a graph $\mathcal{G}$ is another graph formed from a subset of the vertices and edges of $\mathcal{G}$. The vertex subset must include all endpoints of the edge subset, but may also include additional vertices.

A graph is said to be *connected* when there is a path between every pair of nodes, that is there is a sequence of edges which connects a sequence of every two vertices in the graph.

A *connected component* of an undirected graph is a subgraph in which any two nodes are connected to each other by paths, and which is connected to no additional nodes in the supergraph. Analogously, in the case of directed graphs, a *strongly connected component* is a subgraph in which any two nodes are connected to each other by directed paths, and is maximal with this property.

Note that our definition of graph does not account for self-loops, links from a

node to itself, nor multiple edges, pairs of nodes connected by more than one link.

Network metrics Topological properties of networks can be described at multiple scales: local-scale refers to node-level, meso-scale refers to groups of nodes, global-scale involves the whole graph.

At a local-scale, properties are often referred to as centrality measures, which identify the 'most important' nodes in the network. One example is the *degree*, the number of edges incident to the node. The analogous of degree for weighted networks is called *strength*, it is computed as the sum of the weights of the edges incident to a certain node. Nodes with high degree or strength within a graph are called *hubs*.

At a meso-scale, it is possible for instance to study the tendency of the nodes in the network to be grouped into potentially overlapping sets of nodes such that each set is densely connected internally. *Modules* or *communities* are groups of nodes presenting dense connections within groups and sparser connections between groups.

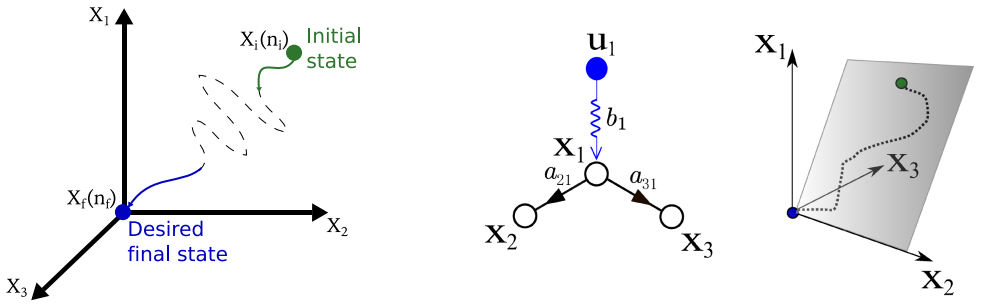
Global scale properties usually study the efficiency of information transfer within the network. For instance, global-efficiency represents a measure of integration, and local-efficiency a measure of segregation [57, 63].

3.2 Controllability theory

Mathematical control theory is the area of application-oriented mathematics that deals with the basic principles underlying the analysis and design of control systems. To control an object means to influence its behavior so as to achieve a desired goal, the central tool is the use of feedback in order to correct for deviations from the desired behavior [156]. Controllability is a prerequisite for control.

The controllability property of a system refers to the fact that any initial state can be steered to any final state by choosing the inputs appropriately, the intuition behind this idea is represented in **Figure 3.1**.

Controllability of a system [96] A *state determined dynamic system* defines, for every pair of initial and final times, n_0 and n_1, a mapping of the initial state $x(n_0)$ and an input trajectory $u(n)$, $n \in [n_0, n_1]$, to the state at time n_1, $x(n_1)$. We shall use the notation $s\left(n_1, n_0, x(n_0), u(\cdot)\right)$ to denote this *state transition mapping*, i.e. $x(n_1) = s\left(n_1, n_0, x(n_0), u(\cdot)\right)$ if the initial state at time n_0 is $x(n_0)$, and the input trajectory is given by $u(\cdot)$.
Consider a state determined dynamical system with a transition map $s\left(n_1, n_0, x_0, u(\cdot)\right)$



(a) Initial and target state.

(b) External input on a node, [102].

Figure 3.1: **The intuition behind the controllability property of a system** refers to the fact that (a) any initial state can be steered to any final state by choosing the input appropriately and (b) applying an external input to a node (here X_1), it is possible to constrain the state of the whole network.

such that

$$x(n_1) = s\left(n_1, n_0, x(n_0), u(\cdot)\right) \tag{3.1}$$

The dynamical system is said *controllable on* $[n_0, n_1]$ if

$$\forall x_0, x_1,\ \exists u(\cdot) \in \mathcal{U} \text{ such that } s\left(n_1, n_0, x_0, u(\cdot)\right) = x_1 \tag{3.2}$$

The dynamical system is said *controllable at* n_0 if

$$\forall x_0, x_1,\ \exists n_1 \geq n_0 \text{ and } u(\cdot) \in \mathcal{U} \text{ such that } s\left(n_1, n_0, x_0, u(\cdot)\right) = x_1 \tag{3.3}$$

Continuous time linear system Consider the *continuous time linear system* with $n \in \mathbb{R}$

$$\dot{\mathbf{x}}(n) = A(n)\mathbf{x}(n) + B(n)\mathbf{u}(n) \tag{3.4}$$

We study the case in which matrices $A(n)$ and $B(n)$ are independent on time (A, B), called linear time invariant (LTI). **Equation 3.4** becomes

$$\dot{\mathbf{x}}(n) = A\mathbf{x}(n) + B\mathbf{u}(n) \tag{3.5}$$

The vector $\mathbf{x}(n) \in \mathbb{R}^N$ captures the state of a network of N nodes at time n. For example, $\mathbf{x}_i(n)$ can denote the amount of traffic that passes through a node i in a communication network [131], transcription factor concentration in a gene regulatory network [92], the magnitude of neurophysiological activity in brain networks [65], or the gene expression in a gene coexpression network.

The $N \times N$ matrix A is the state matrix which describes the structure of the system. It is the adjacency matrix of the network and describes the system wiring

diagram and the interaction strength between the components, for example the traffic on individual communication links, the strength of a regulatory interaction, the number of white matter streamlines connecting brain regions, or the significant correlations among gene expressions.

Finally, B is the $N \times M$ input matrix ($M \leq N$) that identifies the nodes controlled by an outside controller. The system is controlled using the time-dependent input vector $\mathbf{u}(n) \in \mathbb{R}^M$ imposed by the controller. Despite the fact that in general the same signal $\mathbf{u}_i(n)$ can drive multiple nodes, we will consider only the case in which each signal controls exactly one node. This hypothesis leads to a matrix B such that each row and column contains only one nonzero element, corresponding to the node that will receive the external input $\mathbf{u}(n)$.

Nodes identified by matrix B when the system is controllable are called *driver nodes*. Notice that there may be multiple sets of driver nodes able to achieve control over the system.

If we apply the external input to a single driver node, $M = 1$, the matrix B becomes a column vector, and we call the problem *single-input*.

The single-input setting allows us to exploit the definition of *control centrality* as in Liu et al. [104]. Mathematically, control centrality of node i captures the dimension of the controllable subspace or the size of the controllable subsystem when we control node i only.

Note that to identify and analyze the control input $\mathbf{u}(n)$ is outside the scope of this work.

We further note that the somewhat odd notation of using n to address a continuous time is motivated by the choice of reserving t to indicate the targets in the rest of the thesis.

3.2.1 Controllability criteria for LTI systems

Classically, there are three main equivalent criteria to test for controllability in a LTI system. Those are the Kalman criterion [50, 82], the Popov-Belevitch-Hautus criterion [24, 156], and the Gramian criterion [25].

Kalman criterion

The Kalman criterion, also known as the Kalman rank condition, is a necessary and sufficient condition for the controllability of a system [50, 82].

Consider a LTI system as in **Equation 3.5**, then define the $N \times NM$ *controllability matrix* Q as

$$Q = \begin{bmatrix} B & AB & A^2B & \cdots & A^{N-1}B \end{bmatrix} \tag{3.6}$$

We say that a system is controllable if the controllability matrix has full row rank

$$\text{rank}(Q) = N \tag{3.7}$$

That is, if the system is controllable, Q will have N columns that are linearly independent, then every state is reachable by giving the system proper inputs. This criterion can be implemented for any kind of graph: binary or weighted, directed or symmetric.

Trivially, when all nodes receive an external input, B becomes the identity matrix of order N, thus Q is always full rank and the system is controllable.

In the Kalman rank condition, for any $N' > N$,

$$\text{rank}\begin{bmatrix} B & AB & \cdots & A^{N'-1}B \end{bmatrix} = \text{rank}\begin{bmatrix} B & AB & \cdots & A^{N-1}B \end{bmatrix} = \text{rank}(Q) \tag{3.8}$$

Thus, only powers of A up to N are required to test for controllability.

We notice that if the weights in the adjacency matrix are all equal, for instance all ones, this may be seen as a worst-case scenario for controllability. In fact, it is more likely for two rows to be linearly dependent compared to the case with real weights.

There are some limitations to the implementation of the Kalman rank condition. First of all - as the two other equivalent criteria - it is a brute-force approach, that has exponential complexity if requested to test all possible numbers of driver nodes and binomial complexity for a fixed number of driver nodes.

In fact, for a fixed number of drivers M, $1 \leq M \leq N$, there are $\binom{N}{M}$ different configurations of the matrix B, in fact we choose M ones representing the drivers among the N columns, and the complexity can be computed as $\sum_{m=1}^{N} \binom{N}{m} = 2^N - 1$.

Along with the exponential complexity in time, the problem of executing an accurate rank test is ill-conditioned and it may be sensitive to roundoff errors and uncertainties in the matrix elements [102].

Interpretation of the controllability matrix in the single-input case The controllability matrix Q is made of a linear combination of the column vector B and powers of A. For a binary adjacency matrix, the element $[A^k]_{ij}$ can be interpreted

as the number of walks of length k between each pair of nodes, from node j to node i (see [124], p.136).

Then, in the single-input case, the interpretation of the elements Q_{ij} is as follows. The first column Q_{i1} corresponds to the nodes receiving the input B, then elements on the second column Q_{i2} represents the number of paths of length 1 from the driver to the other nodes i. In general, Q_{ij} is the number of paths of length $j-1$ from the driver to the other nodes in the network.

Popov-Belevitch-Hautus (PBH) criterion

This criterion presents a condition in terms of the eigenmodes of the system [24, 156].

Consider a LTI system, as in **Equation 3.5**, then the system is controllable if and only if

$$\lambda \in \mathbb{C},\, z \in \mathbb{C}^N, z^*A = \lambda z^*,\, z^*B \geq 0 \Rightarrow z = 0 \tag{3.9}$$

It is also possible to express the PBH test saying that the system is controllable if and only if it holds

$$\text{rank}\,(\lambda I_N - A,\, B) = N \quad \forall \lambda \in \mathbb{C} \tag{3.10}$$

where I_N is the identity matrix of order N.

Note that rank $(\lambda I_N - A,\, B)$ is less than N only if λ is an eigenvalue of A.

This is also equivalent to say that there is no left eigenvector of A that is a left zero vector of B.

Gramian criterion

The Gramian controllability criterion states that a system is controllable if and only if its *controllability Gramian* is full rank [25, 96].

An alternative representation of the continuous system in **Equation 3.4** is given by

$$\mathbf{x}(n_1) = \phi(n_1, n_0)x_{n_0} + \int_{n_0}^{n_1} \phi(n_1, n)B(n)u(n)dn \tag{3.11}$$

Then, the *controllability Gramian* on the time interval $[n_0, n_1]$ for a LTI continuous system is defined as

$$\mathbf{W}_{[n_0,n_1]} = \int_{n_0}^{n_1} e^{A(n_0-n)}BB^T e^{A^T(n_0-n)}dn \tag{3.12}$$

where B^T and A^T stand for the transposed matrices of B and A respectively.

The system is controllable if its controllability Gramian is full rank on the time interval $[n_0, n_1]$.

In discrete time, the *controllability Gramian* on the time interval $[n_0, n_1]$ of a LTI system is defined as

$$\mathbf{W}_{[n_0,n_1]} = \sum_{j=n_0}^{n_1-1} A^j B B^T (A^T)^j \tag{3.13}$$

Analogously, the system is controllable if its controllability Gramian is full rank on the time interval $[n_0, n_1]$.

3.2.2 Structural controllability

Structural controllability was introduced to give more emphasis to the topology of networked systems and to overcome known issues of classical controllability criteria, such as the difficulty of parameter estimation and the occurrence of numerical errors in the rank tests computations.

The objects of study of structural controllability are *structured matrices*, in which it is assumed that the elements are either zeros, or independent parameters. These matrices describe binary networks.

Let A and B be structured matrices with a LTI dynamics (**Equation 3.5**). Then we say that the system described by (A, B) is *structurally controllable* if it is possible to choose the non-zero weights in A and B such that the system satisfies the Kalman rank condition. A structurally controllable system can be shown to be controllable for almost all weight combinations, except for some pathological cases with zero measure [32, 98].

The main advantage of structural controllability is that necessary and sufficient conditions for controllability can be provided as graph-theoretical conditions [33, 47, 172].

Maximum cardinality matching One recent important result from Liu et al. [103] was to map the Kalman rank condition (**Equation 3.7**) into the maximum cardinality matching over a graph.

A *matching* in a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a subset $\mathcal{M}$ of the edges $\mathcal{E}$ such that no two edges in $\mathcal{M}$ share a common end node.

A *maximum cardinality matching* is a matching that contains the largest possible number of edges.

A *perfect matching* occurs when all nodes are matched in a symmetric graph, or all but one nodes are matched in a directed graph.

A node is *matched* if it is an endpoint of one of the edges in the matching $\mathcal{M}$, otherwise the node is *unmatched*. In particular, a matched edge in an undirected graph, will produce two matched nodes, both nodes to which it is incident. In a directed graph a matched edge will produce only one matched node.

In the framework of Liu et al. [103], given a maximum cardinality matching, the minimum number of inputs required to achieve control over the system is one if there is a perfect matching, and is equivalent to the number of unmatched nodes otherwise. Then, the driver nodes correspond to the unmatched nodes, or to any single node in the case of perfect matching.

Instead of a brute-force test for controllability, this framework provides us with a method to identify a configuration of driver nodes. The maximum cardinality matching is a NP-hard problem if required to find all the feasible configurations, but since directed graphs can be mapped in symmetric bipartite graphs, heuristics exist to find one configuration in polynomial time [76].

Limitations One limitation of this approach is that, by definition, some categories of matrices cannot be considered structured because their non-zero parameters are not independent. Recent studies are working to overcome this issue in the case of symmetric matrices [95, 113].

In the structural controllability framework it is not possible to give an estimate of the control energy required to achieve a certain task. In fact the answer to the controllability test is binary: either a system is controllable or not, and there are no cues to the difficulty of the problem.

In general, in the structural controllability framework (as in [103]), it is assumed that the same input can be applied to multiple driver nodes. However this is not always a realistic assumption, and it may not be feasible in some practical cases [32].

With these limitations in mind, we developed an exact method to study the role of single driver nodes in controlling a targeted system.

Chapter 4

Step-wise target controllability

4.1 Target controllability

The controllability property of a system refers to the ability to affect the state of the whole network. In practice however, we may be interested in modifying the states of a selected subset of nodes, the *target nodes*. If our goal is to control a subsystem, it is possible to accommodate this constraint in the Kalman criterion, following the framework of [58], and the recent application in [187].

Let $\mathcal{T}$ be the *target set*, the subset of nodes to be controlled, of cardinality $|\mathcal{T}| = S$. We introduce the $S \times N$ matrix C whose rows are rows of the identity matrix of order S, with the non-zero element in correspondence with a target node. Then, we can define the analogous of **Equation 3.6**, the target controllability matrix

$$Q_{\mathcal{T}} = C \times Q = \begin{bmatrix} CB & CAB & CA^2B & \cdots & CA^{N-1}B \end{bmatrix} \tag{4.1}$$

And we can update the Kalman rank condition, **Equation 3.7**. The driver nodes identified in B can control the target set $\mathcal{T}$ if and only if

$$\text{rank}(Q_{\mathcal{T}}) = S \tag{4.2}$$

Note that when $\mathcal{T}$ contains only one target node, $S = 1$, then the target is controllable only if it is reachable from the driver nodes. In fact the matrix C will have one row, and since $\text{rank}(C \times Q)$ takes into account only the row of Q corresponding to the target node, there would be at least a non-zero element only if there is at least one path from the drivers to the target.

Interpretation of the controllability matrix in the single-input target case In the controllability matrix computation of the single-input problem, B filters out

columns of A^k in such a way that only those corresponding to the driver node are retained, **Section 3.2.1**. In addition, in the single-input target problem, the matrix C filters out rows of Q in such a way that only those corresponding to target nodes are retained. As a result, for a binary adjacency matrix, elements of $Q_{\mathcal{T}} = C \times Q$ represent the number of walks starting from the driver (due to matrix B) and arriving in the targets (due to matrix C).

4.2 Step-wise target controllability

Linear systems controllability theory has a wide range of applications and a high versatility. It mainly deals with the problem of identifying the nodes in which to apply suitable inputs such that a system can be steered from an initial state to a desired final state. It can be applied to biological systems, where it is often the case that additional constraints need to be enforced. In particular, a context-specific subsystem may be of interest, and it may not be feasible to simultaneously act on more than one input. These two additional constraints lead to what's called a single-input target control problem.

We addressed this specific problem, with the aim to gain insight from the application of controllability theory to biological systems, and in particular to study what is the role of a single driver node in controlling a target set.

Notice that computing the rank of the controllability matrix is ill-conditioned for a large network, thus it was not feasible for us to test all target nodes at the same time. To overcome this issue, we devised ways to lower the combinatorial complexity of the problem, working on smaller subsystems.

Given a target set $\mathcal{T}$, and a single driver node, we imposed an ordering to the elements of $\mathcal{T}$, and we developed a strategy to identify the highest-ranked controllable configuration of targets. We call this procedure *step-wise target controllability*. This strategy implements the Kalman rank condition in an original way, removing the nodes known to be irrelevant to the constrained problem.

The ordering of the target set is required to lower the combinatorial complexity of the problem. In fact, in general the Kalman criterion returns the number of controllable targets, and there may be up to a factorial number of a priori equivalent controllable configurations of targets.

If $|\mathcal{T}| = S$ target nodes are of interest, but only $\text{rank}(Q_{\mathcal{T}}) = S_1$ of them are controllable, a priori there is a factorial number, $\binom{S}{S_1}$, of possible controllable configurations.

Instead, the ordering of the target set allows us to pinpoint the unique configu-

ration with the highest-ranked targets.

It is trivial to notice that there is a single controllable configuration when all target nodes can be controlled.

Remove non necessary nodes from the computation

Consider a single-input target control problem, in which there is one driver node (B is a column vector) and a set of target nodes $\mathcal{T}$. Because of the interpretation of the target controllability matrix $Q_{\mathcal{T}}$, not all the nodes in the network are required for the computation of the rank condition.

In particular, *i)* because of the presence of B, only nodes accessible from the driver are required, and *ii)* among those, only nodes that are on a path from the driver to the targets need to be considered.

Rows and column not corresponding to the above cases will provide all-zero contribution to the target controllability matrix $Q_{\mathcal{T}}$, because there are no paths that from the driver reach those intermediate nodes and then the targets. Thus, they would not contribute to the rank and can safely be discarded from the computation. Removing those nodes lowers the occurrence of numerical errors because manipulating a sub-system of $N_1 < N$ nodes only requires to compute powers of A up to $N_1 - 1$ to form the controllability matrix, **Equation 3.6**.

4.2.1 The algorithm

The step-wise target controllability strategy mainly takes advantage of two key-concepts: the ordering of the target set, and the fact that a controllable configuration of target nodes can trivially be identified when all targets are controllable. The strategy can be described by a simple algorithm.

Given an ordered target set $\mathcal{T}$, each node in the network is tested as a driver. We compute the *step-wise target control centrality (τ)* of a driver as the number of targets that it can control.

Let $[t_1, ..., t_S]$ be the ordered sequence of target nodes, $\mathcal{T}'$ is the current target set and contains the target nodes we are currently testing, τ^* is the number of controllable targets among the ones already tested. We compute the step-wise target control centrality as follows.

- *Step 0*: the target set contains the first target node, $\mathcal{T}' = \{t_1\}$, the initial step-wise target control centrality is zero, $\tau^* = 0$.
- *Step 1*: build the subgraph $\mathcal{G}'$, made of nodes accessible from the driver that can reach the nodes in the target set $\mathcal{T}'$.

- *Step 2*: perform the Kalman criterion to check for target controllability of the subgraph $\mathcal{G}'$.
  - If the configuration is not controllable, discard the last target added to the target set and test the successive one. At a general step when the last node added to $\mathcal{T}'$ was t_j, it means $\mathcal{T}' = \mathcal{T}' \setminus \{t_j\}$ and $\mathcal{T}' = \mathcal{T}' \cup \{t_{j+1}\}$.
  - Else, include the successive target in the target set and increase by one the step-wise target control centrality. At a general step when the last node added to $\mathcal{T}'$ was t_j, it means $\mathcal{T}' = \mathcal{T}' \cup \{t_{j+1}\}$ and $\tau^* = \tau^* + 1$.
- Repeat steps 1 and 2 until the last target node is tested.

At the end of this procedure, the step-wise target control centrality of the driver currently being tested is $\tau = \tau^*$, and $\mathcal{T}'$ contains its highest-ranked controllable configuration of targets.

The Matlab code associated to the step-wise target controllability is freely available at `https://github.com/BCI-NET/Public`

Toy example

Functioning of the step-wise target controllability framework can be appreciated in the following example. Let us consider a directed full binary tree with $h = 6$ levels, with the root node as the candidate driver. Without loss of generality, we randomly position a target in each level and we rank them according to their height in the tree. Then, we introduce a simple cycle among the first three nodes of the tree, as in **Figure 4.1**. By construction, this configuration is controllable and the entire target set can be fully driven by the driver [104]. However, when computing the rank with Matlab (`https://www.mathworks.com/`) `rank()` function considering the entire network, the returned rank is deficient. Instead, by removing the part of the network that is irrelevant for controllability, the rank is full and we retrieve the entire controllable configuration, even in the case of larger networks, i.e. up to $h = 10$ levels.

Computational complexity

The main advantage in terms of time of our method consists in the fact that we identify a controllable driver-targets configuration by performing j tests when j out of S target nodes are controllable. This is opposed to a brute-force approach that requires to test all the possible configurations with j targets and that, in the worst-case scenario, leads to $\binom{S}{j}$ tests.

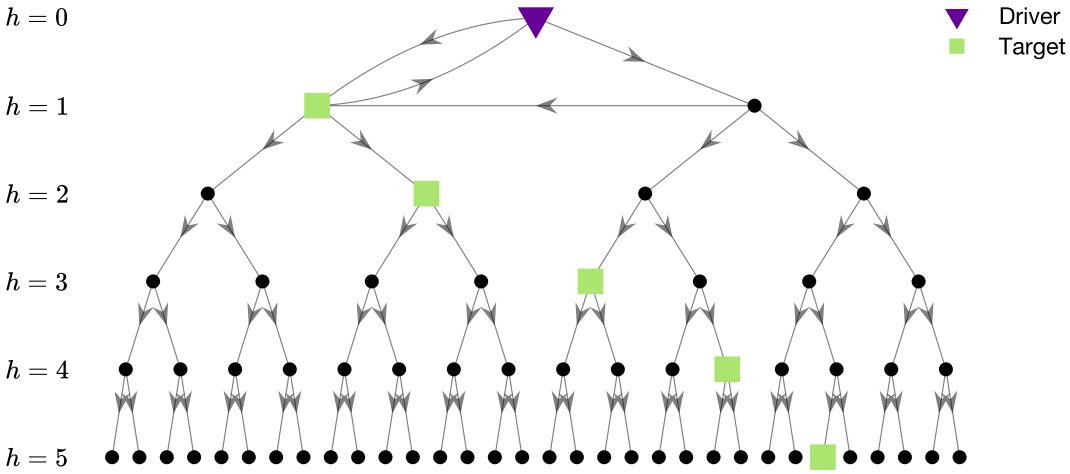


Figure 4.1: **Methodological validation of step-wise target controllability.** We start from a simple directed full binary tree, and we add a cycle among the first three nodes. The driver is the root of the tree, while we put a target node in each level of the tree. Target nodes are then ranked according to their height h. This configuration is fully target controllable by construction regardless of the tree's height h. However, for $h = 5$ (i.e. $N = 63$), the standard procedure computing the rank of the full Kalman controllability matrix cannot retrieve all the targets, as the rank computation is deficient due to numerical errors. Instead, by using the step-wise target controllability we can correctly identify the controllable targets up to $h = 10$, i.e. $N = 2047$.

In addition, because of the step-wise procedure, the method does not need to compute the rank of the full controllability matrix and this allows to minimize possible round-off errors at least for relatively small networks.

To explore the limits of our method in terms of computational complexity, we performed a simulation analysis on synthetic random networks, which vary in number of nodes, connection density and target size, **Figure 4.2**.

We generated a sample of 100 connected and directed random networks with randomly chosen targets. For each node we computed the number of controllable targets through the step-wise target controllability, and through the standard procedure that computes the rank of the full controllability matrix. Results show that in general our method is able to retrieve a larger number of controllable targets. More specifically, when the target set contains 5% of the network nodes, results are quite stable across different connection densities. For larger target-set sizes our method works better when the connection density is relatively low (0.02-0.10). It is important also to notice that the computation of the rank starts to fail in correspondence of larger and denser networks (i.e., $N > 180$ and density > 0.20). While these numbers are in general relatively low and impede to scale up to very large networks,

they still represent ranges that are compatible with typical brain networks obtained from neuroimaging data [22, 23, 44].

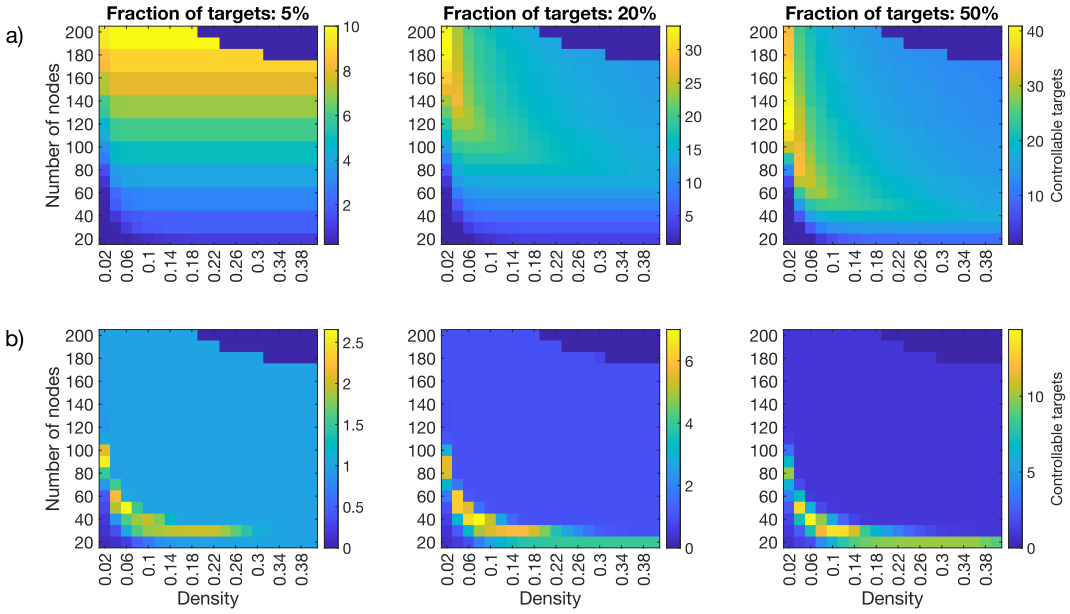


Figure 4.2: **Average number of targets that can be controlled by a single driver node**, computed on samples of 100 connected and directed random networks, at the varying of size and density of the networks, and target set size. Targets are chosen randomly. Panel **a)** corresponds to the step-wise target controllability. Panel **b)** corresponds to a standard procedure that computes the rank of the full Kalman controllability matrix. Results show that in general our method is able to retrieve a larger number of controllable targets as compared to the standard procedure. More specifically, when the target set contains 5% of the network nodes, results are quite stable across different connection densities. For larger target-set sizes our method works better when the connection density is relatively low (0.02-0.10). It is important also to notice that the computation of the rank starts to fail in correspondence of larger and denser networks (i.e., $N > 180$ and density > 0.20).

4.2.2 Limitations and possible developments

The step-wise target controllability framework can be implemented for any network, since the Kalman rank condition on which is based is a general criterion. However, our framework is mainly beneficial to sparse, directed networks. In fact, for this kind of networks the dimensionality reduction has a real impact on the computation. Instead, in the case of undirected graphs it is not possible to find nodes that are not on a walk from the driver to the targets since all nodes belong to the same connected component. Similarly, for directed networks, as the density increases more and more nodes become part of the same strongly connected component, until the whole network is strongly connected.

It may be interesting to implement a similar step-wise procedure varying the

central controllability test. For instance, in some cases the Popov-Belevitch-Hautus criterion may be more numerically stable, or, in the case of a nonlinear dynamics, a feedback vertex set [191] procedure may be implemented.

As for the choice of the target set, also the choice of an ordering is context-specific, and should be considered for each particular application. For instance, targets can be sorted according to their importance in a biological function (e.g. genes) [99, 197] or in discriminating pathological conditions (e.g. brain areas) [106, 151].

It should be noticed that in the step-wise target controllability framework, the procedure does not account for the case in which the ordering of the targets is not unique, two nodes having the same importance. In fact that situation would eventually lead to a factorial number of equivalent configurations, failing to exploit the ordering in the first place, and to successfully lower the complexity of the problem.

To guide in a principled way the choice of an ordering, further studies would be required. For instance, it may be of interest to study the impact of the ordering on the selected controllable configuration of targets in canonical graph models such as Erdös-Rényi random graphs [54] and Barabási-Albert scale free networks [7].

Note on the control input We remark that identifying an appropriate input to apply to the driver nodes in order to achieve a designated control goal is an independent and difficult problem in itself. In general, in order to apply negative feedback to a system, the state matrix A is required to be Hurwitz stable, that is all of its eigenvelues have to be strictly less than zeros [144]. In the cases discussed in this thesis, this is clearly not possible, since we are studying zero-diagonal matrices.

One solution would be to apply a normalization procedure to obtain stable matrices, as in Karrer et al. [84]. We may also choose to run simulations on non-linear models. In this case we would use the LTI framework to identify the drivers, and then well-established non-linear biological models, such as *The Virtual Brain* [149], to study the effect of plausible inputs.

Concerning the applications discussed in this work, it may also be possible to proceed by 'trial and error' in an experimental setting. In fact, in the case of macrophage activation it would be possible to modulate gene expression by means of pharmacological agents, which, however, are only available for a limited number of genes. In brain networks, existing protocols of brain stimulation involving fixed sustained direct currents and trains of monophasic pulses [16] may be studied from a controllability theory perspective.

Due to time-constraints, the problem of finding an appropriate input in the case of the single-input target control problem is outside the scope of this thesis.

4.2.3 Main advantages

Existing methods in general do not allow for target controllability [65, 162], require multiple inputs [88], or provide approximate solutions [40, 58]. Here, we propose an exact method, able to identify the highest-ranking configuration of targets controllable by a single driver node in a reduced amount of time. This method is highly versatile and can be adapted to different applications.

Biological applications

We exploited the step-wise target controllability framework to model two biological processes: the activation of macrophage cells, and aging in the brain.

Aided by experts, we chose context-specific target sets, and their meaningful ordering. Then, we tested each node as a driver, and we obtained a list of their highest-ranked controllable targets.

Differently from most of the literature, we imposed a single driver node, thus we did not require to achieve control over all of the designated targets. Our work is original in the fact that - once found the controllable configurations - it focuses on the study of functional interactions between controllable pairs driver-target, with the aim to uncover relevant biological relations that could be missed by topological analyses.

Chapter 5

Step-wise target control centrality: biological applications

In this chapter are reported text and figures of two recently submitted papers. The papers discuss the two applications in which we tested our method: the macrophage activation process and aging in the brain. Both applications are studied from the perspective of step-wise target controllability, hence there may be some redundancy in the methodological sections.

5.1 Macrophage activation network in multiple sclerosis

Step-wise target controllability identifies dysregulated pathways of macrophage networks in multiple sclerosis.

Giulia Bassignana[1,2], Jennifer Fransson[1], Vincent Henry[1,2], Olivier Colliot[1,2], Violetta Zujovic[1], Fabrizio De Vico Fallani[1,2,*]

1 Sorbonne Université, UPMC Univ Paris 06, Inserm U-1127, CNRS UMR-7225, Institut du Cerveau, Hopital Pitié-Salpêtrière, Paris, France

2 Inria Paris, Aramis Project Team, Paris, France

5.1.1 Abstract

Identifying the nodes able to drive the state of a network is crucial to understand, and eventually control, biological systems. Despite recent advances, such identification remains difficult because of the huge number of equivalent controllable configurations, even in relatively simple networks. Based on the evidence that in many applications it is essential to test the ability of individual nodes to control a specific target subset, we develop a fast and principled method to identify controllable driver-target configurations in sparse and directed networks. We demonstrate our approach on simulated networks and experimental gene networks to to characterize macrophage dysregulation in human subjects with multiple sclerosis.

5.1.2 Introduction

For many biological systems, it is crucial to identify the units, such as genes or neurons, with the potential to influence the rest of the network, as this identification can enable describing, understanding, and eventually controlling the function of the system [19, 79]. Topological descriptors based on network science can indeed be used to quantify such influence in terms of node centrality, such as degree, betweenness, or closeness [124]. However, these descriptors only capture the structural properties of the network and neglect their effect on the dynamics, thus limiting our understanding on the actual influencing power.

Controllability network theory, linking network structure to dynamics through linear or nonlinear models, has been shown to be a more principled approach for identifying driver nodes in an interconnected system [144, 156]. While theoretically these approaches can give a minimum set of driver nodes sufficient to steer the system into desired states, their exhaustive identification might be difficult in practice as there exists in general a very large number of equivalent controllable walks, even in

relatively simple networks [75, 174]. In the case of criteria based on the manipulation of controllability matrices [71, 82], the presence of many walks can for example induce numerical errors due to the different orders of magnitude in the matrix elements.

An alternative solution has recently been proposed to circumvent this limitation, based on the possibility to map the controllability problem onto the maximum cardinality matching over the associated graph [98, 103, 152]. As a result, it is possible to identify a set of driver nodes - at least for directed networks - with linear, and not exponential, time complexity [76]. While this approach elegantly solves numerical issues, it can nevertheless not tell which configuration, among all the possible ones, is the most relevant. In general, there is a factorial number of equivalent configurations (with the same number of inputs) and enumerating all possible matchings [170] rapidly becomes unfeasible, even for simple graphs such as trees [75, 174], bipartite graphs [101], or random graphs [192]. Thus, the research of alternative strategies to characterize the candidate driver nodes is crucial for the concrete application of network controllability tools.

One possibility would be to reduce the original problem into smaller sub-problems under the assumption of specific constraints compatible with the underlying scientific question. On the one hand, for many biological, technological and social systems it is desirable to only control a subset of target nodes (or a subsystem) that is essential for the system mission pertaining to a selected task or function. In this direction some approaches have been recently proposed, based on filtering of the controllability matrix [89] or adaptation of graph matching [39, 58, 95, 194]. However, they do not solve the problem of multiple driver set configurations. On the other hand, technical and experimental constraints often limit the possibility to stimulate many driver nodes in parallel, for example in gene expression modulation [105] or brain stimulation [68]. In these cases, approaches that focus on the ability of single driver node to control the entire network, such as control centrality [104] or single-node controllability [65], do circumvent the multiplicity issue, but can still suffer from numerical errors and approximate results.

To overcome this impasse, we propose an integrated method that combines the advantages of the previous approaches and quantifies the capacity of a single driver node to control a predefined target set. Based on the Kalman controllability condition, our method identifies the part of the target set that can be controlled by a candidate driver. To do so, we introduce a ranking among the target nodes and we iteratively evaluate the controllability of the system by adding one target node at a time in a descending order. This eventually finds a univocal controllable configuration corresponding to the highest ranking. In the following, we first illustrate how our method, named *step-wise target controllability*, works for simple network structures

and we discuss the potential benefits for directed and sparse networks, in terms of space and computational complexity, as compared to alternative approaches. Then, we use it to study molecular networks of macrophage pro-inflammatory activation, derived from ontology-based reconstructions, and identify the driver-target pathway alterations using gene expression data from blood samples of patients affected by multiple sclerosis (MS) and a matched group of healthy controls (HC).

5.1.3 Results

Step-wise target controllability identifies a controllable subset of targets

Let $\mathcal{G}$ be a directed graph (or network) of N nodes (or vertices) and L links (or edges), and $\mathcal{T}$ an arbitrary subset of $S < N$ nodes in the network. The aim is to measure the ability of each node to drive the state of the target set $\mathcal{T}$ from a dynamical system perspective [156]. In the case of linear time-invariant dynamics, the number of controllable target nodes can be obtained computing the rank of the target controllability matrix [35, 58, 121]:

$$Q_{\mathcal{T}} = \begin{bmatrix} CB & CAB & CA^2B & \cdots & CA^{N-1}B \end{bmatrix} \tag{5.1}$$

where A is the adjacency matrix of the network, B is a vector identifying the driver node and C is a matrix selecting the rows of A corresponding to the target, or output, nodes (**Methods**).

A *walk* in the network consists of an alternating sequence of vertices and edges.

The (i, j) entry of $Q_{\mathcal{T}}$ indicates how many walks of length $j - 1$ connect the driver to the target i [15]. Trivially, all the nodes not traversed by these walks do not contribute to the walk lengths and they can be neglected for the purpose of controllability. By removing the irrelevant nodes from the network, A becomes smaller and this results in a target controllability matrix with less columns. Put differently, we avoid the computation of matrix exponentials corresponding to non-existing driver-target walks (**Methods**). In practice, this can be of great advantage for reducing the occurrence of round-off errors during the matrix rank calculations. For example, this is the case for sparse and directed networks, where fewer nodes are reachable as compared to dense and undirected networks.

This can be easily appreciated in the following example. Let us consider a directed full binary tree with $h = 6$ levels, with the root node as the candidate driver. Without loss of generality, we randomly position a target in each level and we rank them according to their height in the tree. Then, we introduce a simple cycle among the first three nodes of the tree (**Figure 4.1**). By construction, this configuration is controllable and the entire target set can be fully driven by the

driver. However, when considering the entire network the returned rank is deficient. Instead, by removing the part of the network that is irrelevant for the controllability, the rank is full and we retrieve the entire controllable configuration, even in the case of larger networks, i.e. up to $h = 10$ levels.

The rank of $Q_{\mathcal{T}}$ gives the number of target nodes $\tau \leq S$ controllable by the driver, but there might be in general many possible equivalent configurations. To overcome this issue, we propose a step-wise procedure that tests the controllability on subproblems of increasing size. First, we introduce a hierarchy among the target nodes and relabel them according to their importance in a descending order, i.e. $t_1 \succ t_2 \succ ... \succ t_S$. Then, we create an empty auxiliary set $\mathcal{T}'$ and we sequentially include the target nodes according to their ordering. At each step, if the rank of $Q_{\mathcal{T}'}$ is full, the new target node is retained, otherwise it is removed from $\mathcal{T}'$. When all the target nodes have been visited, the algorithm returns the set of controllable targets with highest ranking (**Figure 5.1**, **Methods**).

Our method, named *step-wise target controllability*, not only returns for each candidate driver the number of controllable target nodes τ corresponding to the configuration with highest ranking, but also the set $\mathcal{T}'$ of controllable targets.

Driver genes are homogeneously distributed in the macrophage network

To test our method in a biological context, we construct a network representing the interactions between molecules involved in macrophages response to pro-inflammatory stimuli (**Figure 5.2**), with the connections between genes inferred from a previously established network based on literature [139, 143]. This network is of interest in MS due to the chronic inflammation characteristic of the disease, and the generally destructive effects of pro-inflammatory macrophages in MS [17]. Hence, dysregulation of macrophages may lead to aggravated inflammation and disease.

In order to facilitate biological interpretation of the network, we divide the nodes according to molecular function: *sensing*, *signaling*, *transcription* factors or *secreted* molecules (**Table SA1**). We choose the 13 *secreted* molecules as target nodes because they represent the end-products of macrophage pro-inflammatory activation and enable propagation of inflammation to other cells, thus exacerbating chronic inflammation.

To establish a hierarchy among the targets, we use macrophage RNA expression data from a group of MS patients and healthy controls. The macrophages were tested with and without activating stimuli to mimic the pro-inflammatory response. We measure the gene activation as the ratio of the expression between the "pro-inflammatory" and "alert" condition. We then consider the fold change Δ between

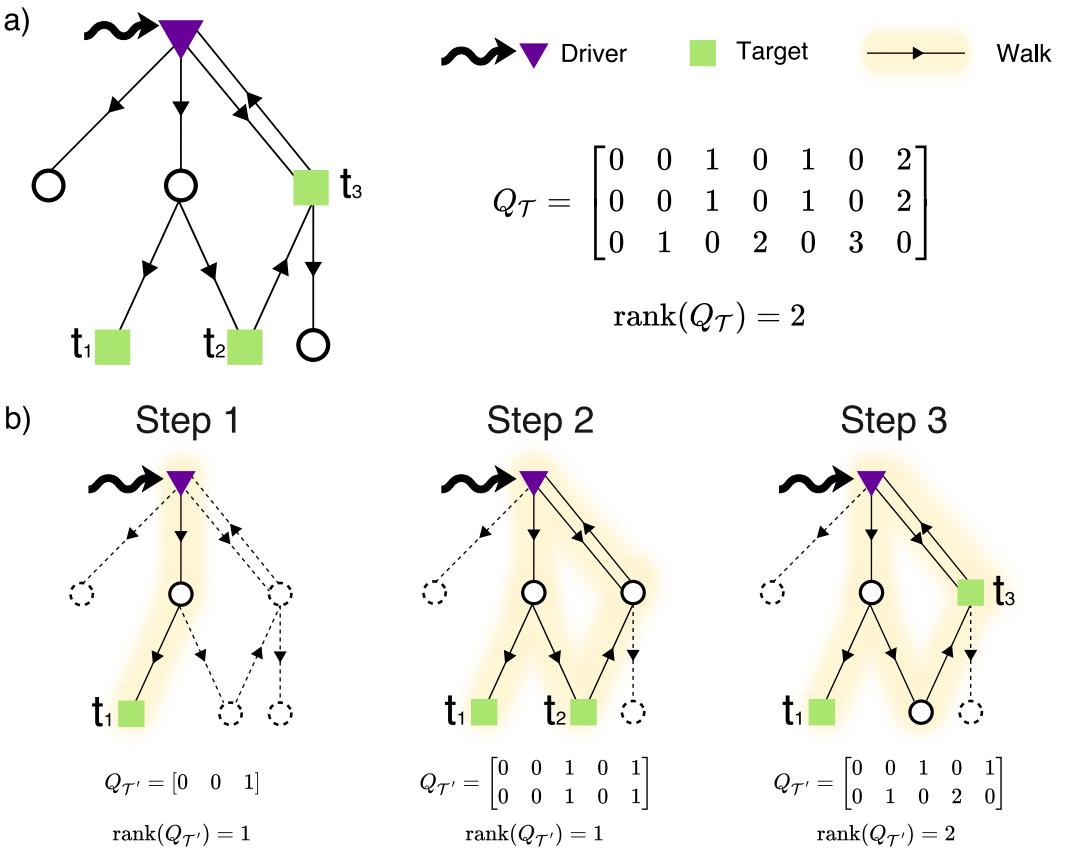


Figure 5.1: **Working principle of step-wise target controllability.** Panel **a)** illustrates a network with one driver and a target set $\mathcal{T} = \{t_1, t_2, t_3\}$ of cardinality $S = 3$. The Kalman condition informs us that only two targets are controllable from the driver, i.e. $\tau = \text{rank}(Q_{\mathcal{T}}) = 2$. However, there might be up to 3 equivalent configurations that are controllable, i.e. $\{t_1, t_2\}$, $\{t_1, t_3\}$, and $\{t_2, t_3\}$. For larger networks, the number of Kalman tests to perform can be prohibitive, i.e. $\binom{S}{\tau}$. Panel **b)**. By introducing a hierarchy among the target nodes, our step-wise method identifies the configuration with the most important nodes by performing only S tests (see **Methods**). In this example, the first step considers the subgraph containing all the walks from the driver to the target set $\mathcal{T}' = \{t_1\}$. The associated controllability matrix has full rank, i.e. $\text{rank}(Q_{\mathcal{T}'}) = 1$. The first target is therefore retained and the algorithm moves to Step 2, by constructing a new subgraph containing the walks from the driver to the target set $\mathcal{T}' = \{t_1, t_2\}$. The rank of the new controllability matrix is now deficient and t_2 is not retained. In Step 3, the new subgraph contains the walks from the driver to $\mathcal{T}' = \{t_1, t_3\}$. Because $\text{rank}(Q_{\mathcal{T}'})$ is full and there are no more targets, the algorithm stops and returns the controllable configuration t_1, t_3.

the gene activation of MS patients and HC subjects (**Methods**). Genes with larger Δ values are ranked first (**Figure SA1**).

Results show that 51% of the tested network nodes can control at least one target (i.e. $\tau > 0$) and that those drivers tend to be homogeneously distributed

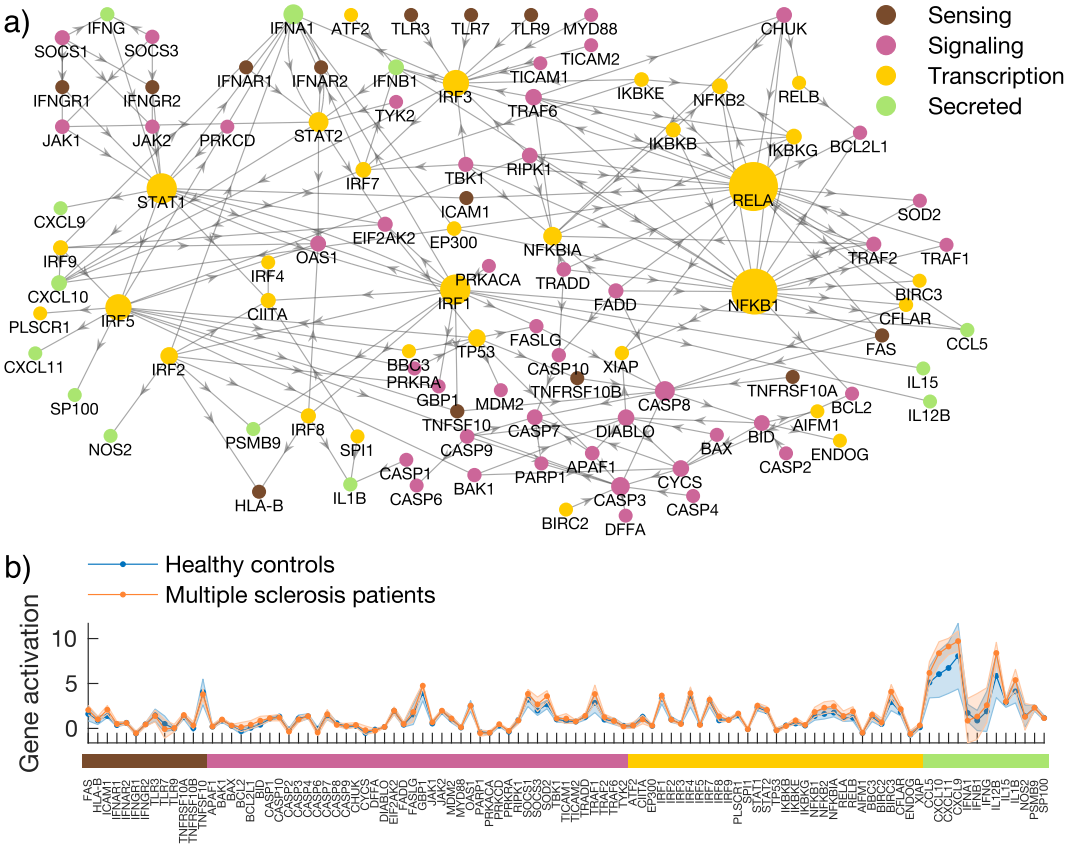


Figure 5.2: **Molecular network and gene activation associated with the pro-inflammatory state of macrophages**. Panel **a)** shows the molecular network reconstructed through ontology-based techniques from the `macrophage.com` repository [139, 143]. The network consists of $N = 101$ nodes corresponding to genes involved in inflammation; for the sake of interpretation, they are organized in four classes, depending on their function in the cell. *Sensing* genes are in the membrane of the cell and start a *signaling* pathway inside the cell, to the *transcription* factors, which promote the production of *secreted* molecules. There are $L = 211$ directed edges representing either activation or inhibition interactions between molecules (**Methods**). The size of the nodes is proportional to their total degree k. Panel **b)** shows gene activation computed as the ratio in expression between the "pro-inflammatory" and "alert" states, based on our RNA sequencing data, generated from monocyte-derived macrophages from blood samples of multiple sclerosis patients ($n = 8$) and healthy controls ($n = 8$) (**Methods**). Solid lines represent group-averaged values, while transparent patches stand for standard deviation.

across classes (**Tables SA2 and SA3**). This indicates a high redundancy in the way the target set can be controlled. Notably, target centrality values are weakly correlated (Spearman rho 0.18, $p < 0.07$) with the corresponding total node degree k, as defined in **Methods**, indicating that the most connected genes (e.g. RELA, NFKB1) are not necessarily the ones that can most efficiently steer the state of the

target set (**Figure 5.3a**).

Almost all of the driver nodes identified by our method can control the target genes CCL5, CXCL10, CXCL11, IFNA1 and IFNB1, which code for inflammatory chemokines and cytokines. This implies a high level of co-regulation among these molecules, with many different actors exerting control over this regulation. Interestingly, the drivers with the highest step-wise target control centrality (τ) values (SOCS1 and SOCS3, $\tau = 10$) belong to feedback systems that control pro- and anti-inflammatory signal transduction by regulating the signaling process triggered in response to IFNγ [111]. In addition, all drivers with $\tau \geqslant 9$ can be seen in our network as a cluster of genes converging onto and including STAT1 (**Figure 5.3a**). This cluster includes the receptors of IFNγ and the signaling molecules responsible for their intracellular effects. This result matches the well-described effects of IFNγ on chemokine production [90, 100] and overall macrophage activation [117].

Robustness of driver nodes to random attacks To assess the stability of our findings to possible errors in the network construction, we performed a robustness analysis simulating different types of alterations to its nodes and links (**Methods**). Results show that removing nodes with higher degree k leads to a greater reduction of step-wise target control centrality in the drivers compared to the removal of low-degree nodes or random removal of nodes (**Figure 5.3b**). For example, by attacking 10% of the nodes we lose 5% of the drivers in the latter cases, while we lose 20% of the drivers when removing the most connected ones (**Figure 5.3b**). This result confirms the crucial role of hubs in biological networks in terms of resilience to random attacks [8] and controllability [134].

When perturbing links, the worst condition is given by their random removal. By attacking 10% of the links around 5% of the drivers are lost. This is intuitively due to the interruption of driver-target walks and to consequent impossibility to control a node that cannot be reached. While randomly rewiring the links has an intermediate impact, adding new links has no effect on the step-wise target control centrality of the drivers (**Figure 5.3c)**. This is of great advantage as it shows that our results will not change if new connections are established or provided by the literature.

Gene dysregulation and altered driver-target coactivation in multiple sclerosis

Using step-wise target controllability, we detect potential directed interactions in the macrophage activation network, but we cannot quantify how changes in the driver's state affect those in the targets. To measure driver-target functional interactions, we

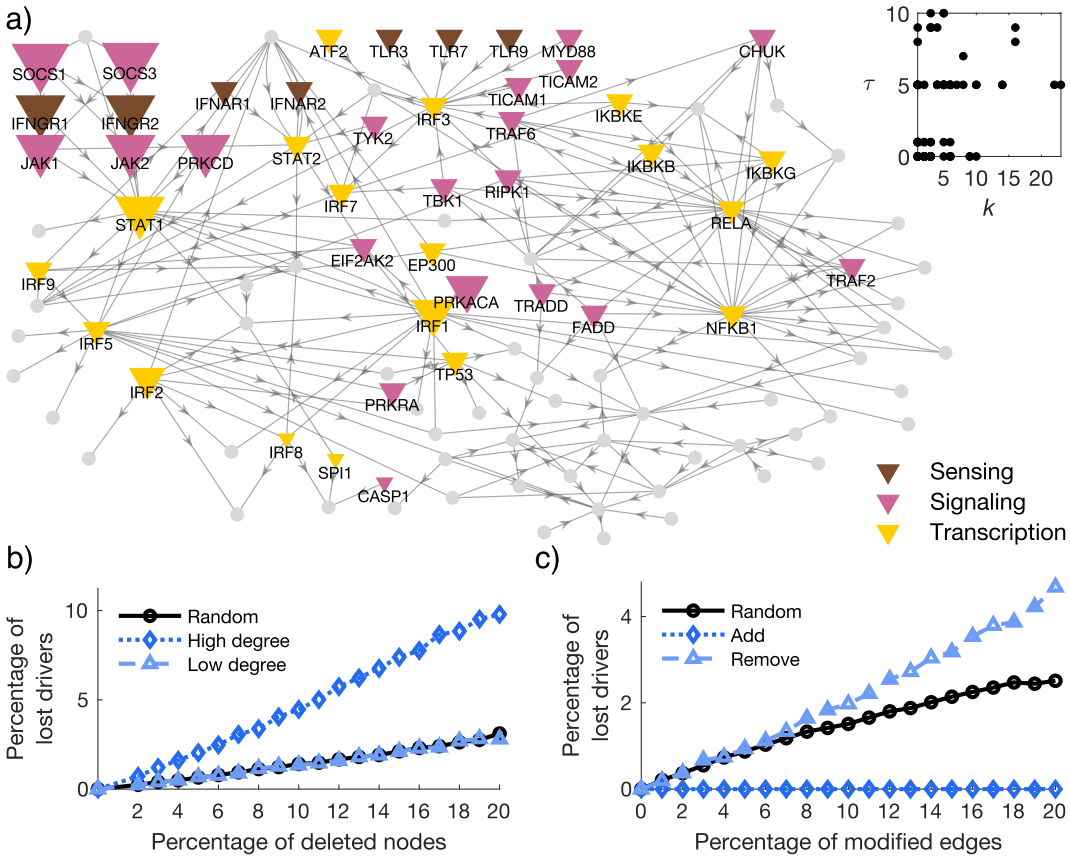


Figure 5.3: **Gene network step-wise target control centrality and analysis of robustness for the driver nodes.** In panel **a)** the size of the nodes codes the step-wise target control centrality (τ) values. Nodes with $\tau = 0$ are classified as not-drivers and are represented in gray. The inset shows that τ values cannot be merely predicted by node degree k (Spearman rho 0.18, p < 0.07). Panel **b)** shows the percentage of driver nodes ($\tau > 0$) that are lost when removing nodes in a random fashion (black circles), or preferentially attacking high-degree (blue diamonds) or low-degree nodes (light blue triangles). Panel **c)** shows the percentage of driver nodes that are lost when randomly rewiring (black circles), adding (blue diamonds) or removing edges (light blue triangles).

compute the Spearman correlation between the gene activation of controllable driver-target pairs, for the HC and MS groups (**Methods, Tables SA4 to SA7**). We call *coactivated* the genes exhibiting a significant correlation (p$<$ 0.05). Results show that in general only a moderate fraction (21%) of all the possible driver-target genes are coactivated (**Figure 5.4a, Tables SA4 to SA7**). For both HC and MS groups these interactions tend to primarily involve signaling functions (**Figure 5.4b**). However, the number of driver-target coactivations is lower in the pathological condition ($MS = 19$ versus $HC = 36$). More importantly, they differ from those observed in the HC group (**Figure 5.4b**). This is particularly evident for target IFNA1, which

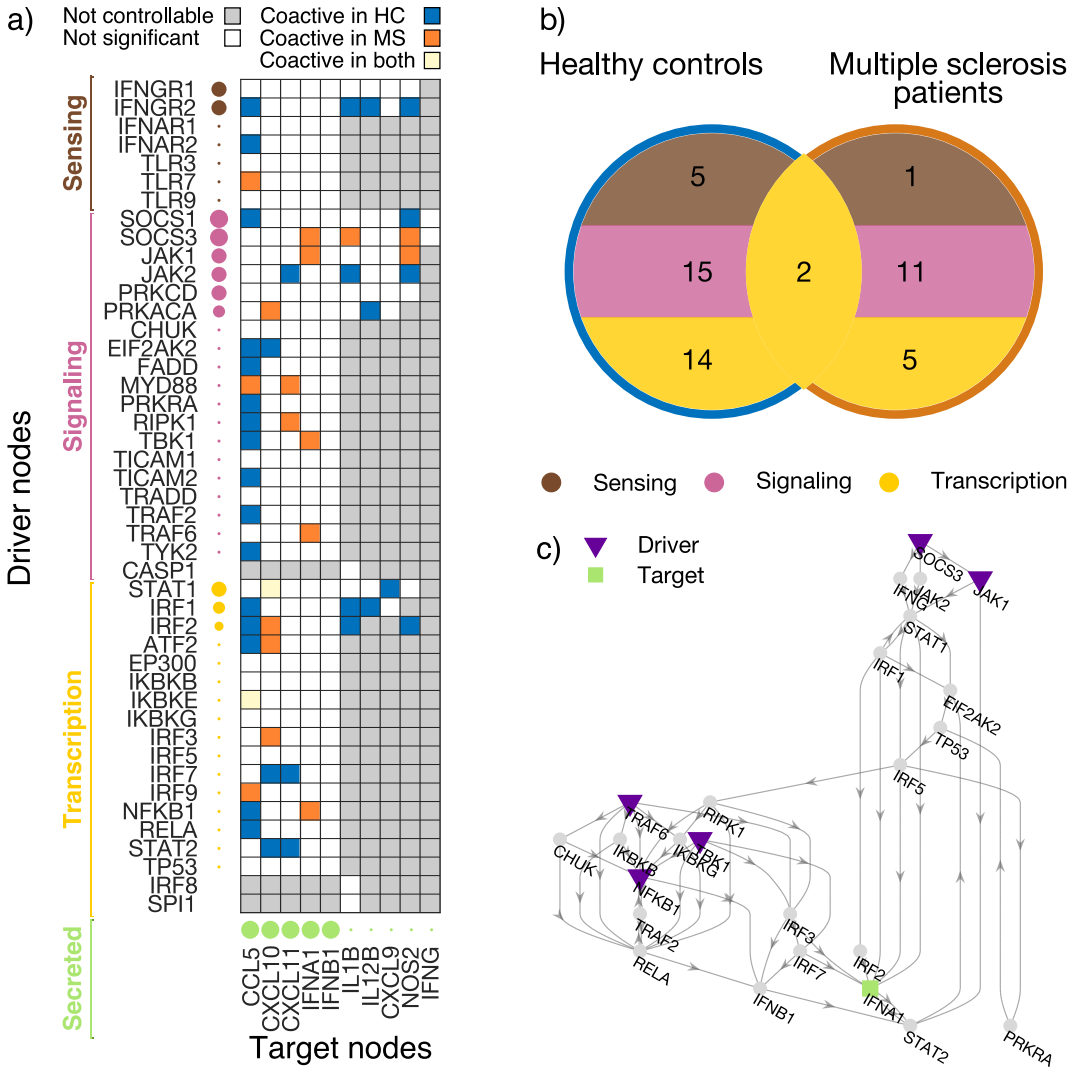


Figure 5.4: **Altered driver-target coactivation in multiple sclerosis.** Panel **a)** reports the coactive driver-target pairs, computed as significant Spearman correlations ($p < 0.05$) between the gene activation of controllable driver-target pairs, for the healthy control (HC) (blue squares) and the multiple sclerosis (MS) (red squares) groups. White squares indicate that there is a controllable walk from the driver to the target, but that their correlation is not significant. Grey squares mean that there is no controllable walk for driver-target pairs. The size of the circles for driver nodes codes for their step-wise target control centrality (τ) values. For target genes, circle sizes represent the number of driver nodes that can control them. Panel **b)** Venn diagram showing a decrease in number of driver-target coactivations in the MS patients as compared to HC. In both groups, these functional interactions tend to predominantly involve signaling genes. Panel **c)** subnetwork of the walks from all the drivers coactivated with the target IFNA1.

only exhibits coactivations with signaling and transcription drivers in the MS group

(**Figure 5.4c**).

Because the macrophage network edges are fixed and reconstructed from known protein-protein interactions, differences in coactivation can be essentially attributed to altered regulation of transcription. Hence, our hypothesis is that the observed functional reorganization can be explained by the dysregulation of specific genes along the controllable walks from the drivers to targets. To test this prediction, we examine all the pairs of genes whose coactivation appears or disappears in the MS group (**Figure 5.4a**). We found that 47/51 of these differentially coactivated pairs present at least one dysregulated gene (i.e., fold change $|\Delta|$ above the 75-percentile, **Methods**) on the walk from the driver to the target (**Figure 5.5, Tables SA8 to SA12**).

We find in total 14 dysregulated nodes on any of these walks. The genes that most frequently appear are NFKB1, IFNA1 and IFNB1 (36/51 walks). They are present on all walks that end with targets CCL5, CXCL10, CXCL11, IFNA1 or IFNB1, i.e., the 5 targets that could be controlled by most drivers. This points to their dysregulation being a potent disruptor of the normal network functioning. The co-occurrence of these three dysregulated genes can be explained by a feedback loop in which NFKB1 activates IFNB1, and IFNA1 and IFBN1 both activate STAT2, which through several intermediates can influence all three genes (**Figure SA2**). Indeed, this stems from the fact that all these nodes belong to the main connected component of the network, i.e. a subnetwork in which every node is reachable from any other node.

Taken together, these results indicate that the aberrant reorganization of functional interactions in the MS group is associated with the presence of dysregulated genes along the controllable walks of the macrophage network.

Switch of SOCS-gene coactive drivers reflects dysregulated inflammatory response Because drivers are crucial for steering the target network's state, we focus on the subnetwork specifically involving the dysregulated drivers (IRF8, NFKB1, SOCS1, SOCS3, TLR7) and the walks towards the respective controllable targets **Figure 5.6**. By looking at how driver and target nodes are differently coactivated in healthy controls and MS patients, we obtain a much clearer description of the gene dysregulation effects. First, many of the previous results can be now appreciated in finer detail, such as i) the reduction in number of coactivated driver-target pairs in MS, ii) the large number of targets that can be controlled by SOCS1 and SOCS3, and iii) the potential of NFKB1, IFNA1 and IFNB1 to affect the driver-target functional interactions.

Second, we report an interesting mechanism involving the drivers with the high-

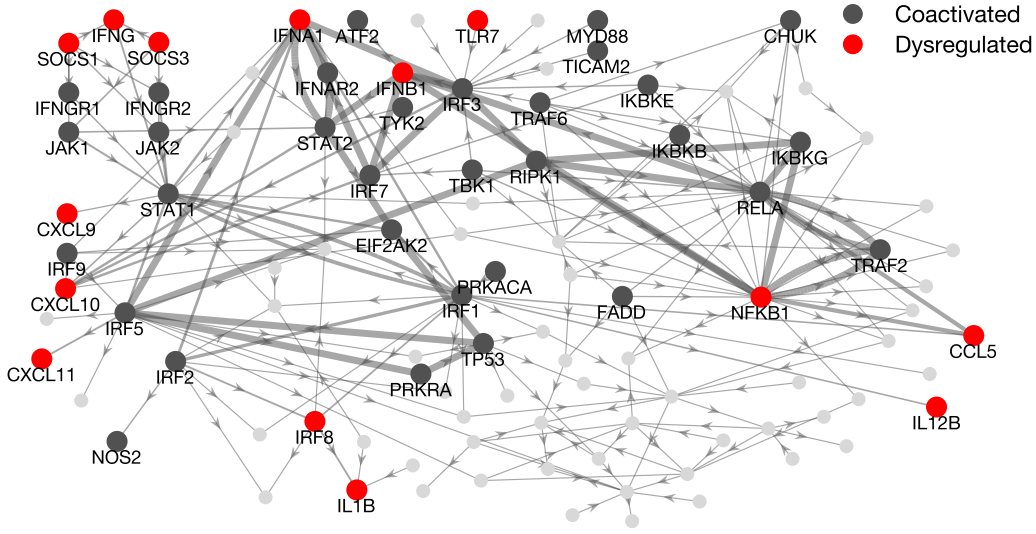


Figure 5.5: **Pooled visualization of dysregulated genes along differentially coactivated driver-target walks.** Highlighted genes indicate all nodes on walks between coactivated driver-target pairs, either in the healthy control group, or in the multiple sclerosis patients group. Dysregulated genes are shown in red. Edge thickness is proportional to the number of times they are traversed by walks connecting a driver to a target node (information not reported here).

est τ centrality values, i.e. SOCS1 and SOCS3. In the HC group, SOCS1 is coactivated with the targets while SOCS3 does not exhibit any significant correlation. In the MS group, we observe the opposite, i.e. SOCS1 is silent while SOCS3 becomes coactive. Because both driver genes are dysregulated, the observed "switch" mechanism could be therefore associated with the altered pro-inflammatory response of the MS group. Indeed, these two molecules are known to be strong modulators of macrophage response: SOCS1 inhibits the signaling of pro-inflammatory genes while SOCS3 is known to be an important actor in inflammatory response, with the ratio of the two proteins determining the actual effect [182].

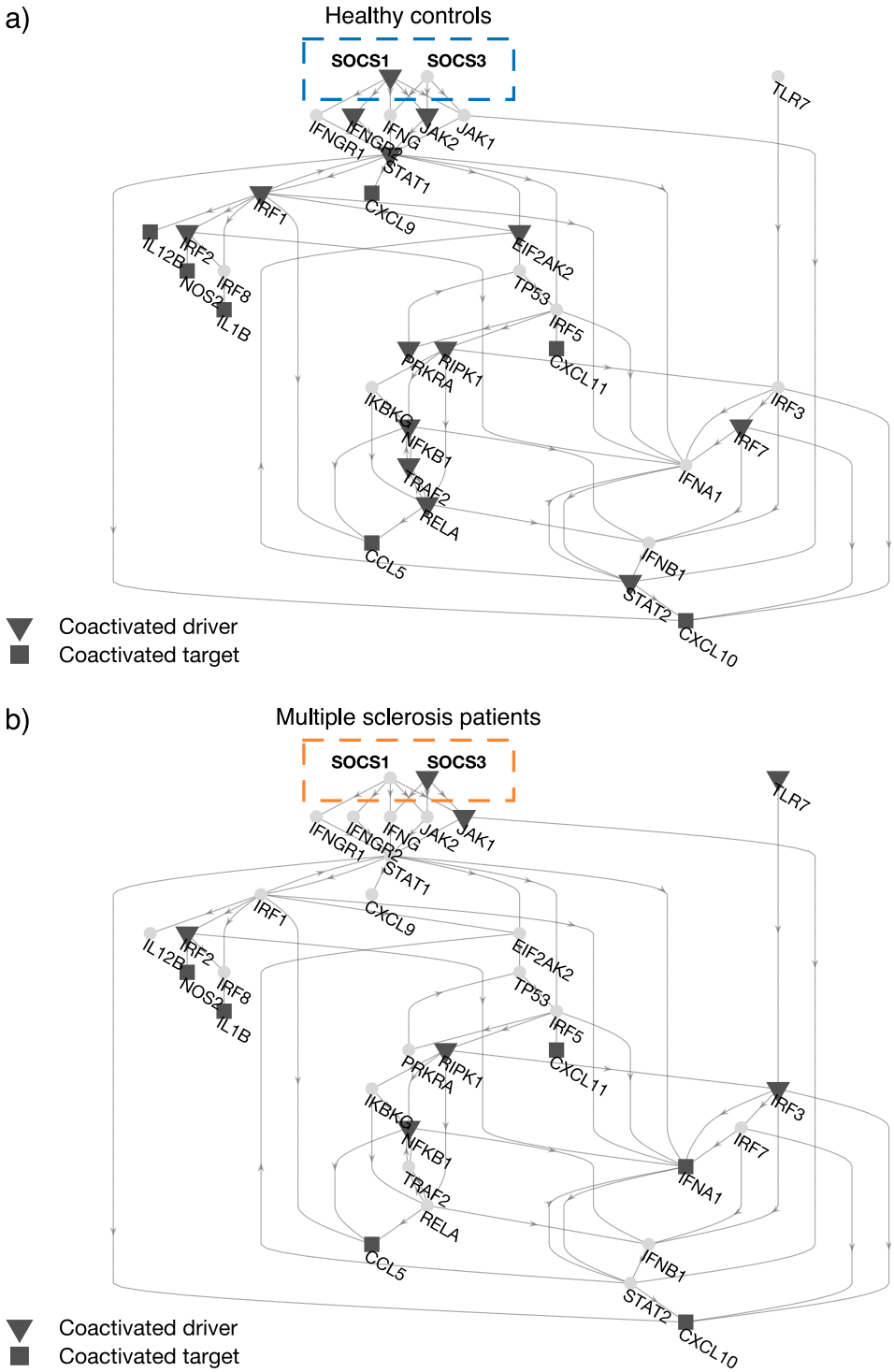


Figure 5.6: **Dyregulated drivers and coactivation switch for SOCS-genes.** The subnetwork includes all dysregulated drivers (IRF8, NFKB1, SOCS1, SOCS3, TLR7) and their controllable targets. Panel **a)** shows coactivated pairs for healthy controls (HC), panel **b)** shows coactivated pairs for multiple sclerosis (MS) patients. A coactivation switch can be appreciated between the HC and MS groups. SOCS1 and SOCS3 are respectively coactive and silent in HC, while they invert their role in the MS group.

5.1.4 Discussion

Identification of controllable configurations in complex networks

Network controllability refers to the ability to drive an interconnected dynamical system from any initial state to any desired final state in finite time, through a suitable selection of inputs [144, 156]. In recent years, an increasing number of research groups from different disciplines have focused their efforts on identifying the minimum set of driver nodes or quantifying the capacity of single nodes to control the entire network, as well as parts of it [14, 51, 66, 102, 109, 114, 119, 161, 163, 175, 185]. Despite being theoretically attractive, network controllability still suffers from computational issues that limit its impact in concrete applications. This is mainly due to the presence of multiple equivalent controllable walks in a network that make the associated controllability problem ill-posed and/or the resulting solution space very big [144, 156].

To reduce such complexity, we propose a method based on control centrality, which was previously designed to quantify the ability of one node to control directed networks [104]. First, we define the *target* control centrality to measure the controllability of a specific part of the network, i.e., a predefined target set. Because edges are directed, this has the advantage to ignore the part of the network that is not traversed by the walks connecting the driver to the target set. Second, we introduce an ordering among the target nodes and perform a step-wise controllability test with increasing size. Because of the ranking, only one controllable configuration will be identified, i.e. the one with the highest ranking (**Figure 5.1**). To test the controllability of the driver-targets configuration at each step we adopted the Kalman criterion [82, 144]. However, the entire iterative framework is quite flexible and other methods, such as Gramian condition [88], Popov-Belevitch-Hautus criterion [95], or feedback vertex set [191], could be used as alternative controllability criteria.

While the step-wise target controllability achieves the identification of one solution in a significantly reduced amount of time, it is important to state that the method for ordering of the targets is a subjective choice. For example, targets can be sorted according to their importance in a biological function (e.g. genes) [99, 197] or in discriminating pathological conditions (e.g. brain areas) [106, 151]. The choice of sorting criteria should reflect the specific scientific question. In a more network-centric approach, ordering could rely on the ranking of topological network properties such as node centrality measures [124]. While this is beyond the scope of our study, we feel that this should be more extensively investigated by future research.

Control pathways in macrophage molecular networks

The study of the molecular interactions is crucial to the understanding of the basic functions of the cell such as proliferation or apoptosis [110, 164]. Determining the connection mechanisms that rule a specific biological function can significantly impact our daily life by providing new therapeutics to counteract diseases [52, 115, 168]. Studying molecular networks is however difficult, because in general we do not know the true functional interactions of a cell and indirect techniques such as gene co-expression are typically employed to infer such connections [171]. Based on correlation analysis, these methods cannot inform on the causal nature of the interactions. More importantly, the reliability of the estimated network critically depends on the number of interactions to number of data samples ratio, which is in practice very low [171].

To overcome this limitation, we reconstruct the directed gene interactions associated with the inflammatory state of the human macrophages by adopting a novel ontology-based approach that integrates the available information from multiple datasets and results in the literature [74]. Previous studies show that the number of driver nodes in biological networks is rather high due to their sparse and heterogeneous nature [103, 145]. Consistently, we find that a large percentage of genes (51%) can control at least one secreted molecule in the target set. Our results also confirm that, despite being crucial for global communication, hubs (e.g. RELA) are not always the most important from a network controllability perspective (**Figure 5.3**a). This stems from the theoretical impossibility to diversify the input signals to all the connected neighbors [103]. The found driver genes are heterogeneously distributed across the tested gene classes. However, our method highlights SOCS1 and SOCS3 as the drivers with the highest step-wise target control centrality, with other IFNγ-response-related genes showing similar values. This is in line with the known effects of SOCS-genes and IFNγ on molecules secreted by pro-inflammatory macrophages [90, 100, 111, 117], supporting the ability of this method to identify biologically relevant drivers.

Overall, these results uncover the existence of potential causal influences from candidate driver genes to the secreted molecules in the human macrophage activation network. Because the identified driver nodes are robust to network alterations, notably when adding new links (**Fig. 5.3**c, blue diamonds), the obtained results are expected to be sufficiently resilient to the integration of new gene-gene interactions. Notably, our results are also relatively robust to the deletion of links, a situation which is often associated with filtering or thresholding out the weakest or less relevant connections in biological networks [55] (**Fig. 5.3c**, light blue triangles).

From a different angle, our approach can be seen as a new way to filter in-

formation in complex networks and focus on the specific nodes (the drivers) or node pairs (driver-target). This might have important consequences when studying genome-wide databases where the high number of elements can make prohibitive the assessment of significant gene expressions and/or co-expressions [13, 120].

Dysregulated genes and aberrant interactions in multiple sclerosis

Multiple sclerosis is an immune-mediated disease in which the immune system erroneously attacks myelin in the central nervous system. There are many neurological symptoms, including motor and cognitive deficits, that can vary in type and severity depending on the attacked central nervous system regions [70]. The role of macrophages in MS is crucial because of their ability to obtain a pro-inflammatory activation state, including the release of pro-inflammatory cytokines and leading to central nervous system tissue damage [29]. Hence, dysregulation of macrophages may lead to autoimmunity and persistent inflammatory diseases [160]. While the etiology of MS is still not well-understood there is a large consensus on its genetic basis and on the importance of unveiling the underlying network mechanisms [3].

In this study, we combined network controllability tools and gene expression data to detect the genes responsible for altering the macrophages action in multiple sclerosis. Differently from standard approaches, where the attention is focused on the identification of the driver nodes in a network, we here propose an alternative way of exploiting network controllability. We first show that the macrophage inflammatory state in the MS group was characterized by a drastic alteration of the coactivations in the driver and target genes (**Figure 5.4**). Such absence of coordination was in general associated with the presence of dysregulated genes along the walks from the driver to the target node. Notably, the pathological dysregulation of NFKB1, IFNB1 and IFNA1, which belong to the same feedback cycle (**Figure SA2**), critically affects several driver-target functional interactions (**Figure 5.5**).

Finally, our approach allows to identify a shift mechanism for dysregulated SOCS1 and SOCS3 drivers, showing opposite coactivation patterns in MS patients compared to the healthy controls (**Figure 5.6**). These results suggest that experimentally stimulating SOCS3 - a strong inducer of pro-inflammatory response - might be more effective for moving the state of the altered secreted molecules towards physiological configurations. Taken together, these results might have practical consequences on how to design intervention strategies and counteract disease phenotype.

Methodological considerations

Our method uses Kalman controllability rank condition [82] to quantify the centrality of the driver nodes. This criterion assumes that the investigated system has a linear dynamics, **Equation 5.2**. In our case, this means that the changes in the gene activation would follow a linear trend. While this is in general not true and difficult to ascertain, it appears that results from non-linear tests are often dominated by linear relationships [155, 159]. Furthermore, a significant fraction of the data analysis and modeling deals exclusively with linear approaches as they are simpler, easy to interpret and serve as a prerequisite of nonlinear behavior [102].

Another peculiarity of our approach is the assumption of time-invariant interactions in the molecular gene network. On the one hand, this assumption allows to better exploit the well-established results and tools in network controllability [103]; on the other hand, it might conflict with existing literature looking for biological connectivity changes between conditions or populations such as differential gene co-expression [28]. Here, we hypothesized that the activation state of each node (in terms of gene expression) could eventually change but not the underlying network structure. Thus, our network - obtained from detailed maps of the macrophage cells - would only act as a substrate/proxy for functional interactions, such as correlated gene activities.

Our method requires a specific ordering of the target nodes. While this can be typically achieved in many biological applications - by ranking nodes according to their state (e.g., gene expression [99, 197], brain activity [106, 151]) - challenges might still remain in general. When it is not possible to impose a ranking of the nodes from external knowledge, another possibility is to derive it from the network structure taking into account, for example, node centrality values. However, there might be multiple nodes with the same centrality value, which would impede a proper ranking. In those situations, multiple node centrality measures could be integrated to get a more heterogeneous distribution (e.g., degree, strength, betweenness, closeness) [124]. Another possibility is to add equally important targets at the same iteration step and test if they are simultaneously controllable. While this procedure is suboptimal and may underestimate the number of controllable targets, it still minimizes the computational complexity related to testing multiple driver-target combinations.

We finally notice that our method is conceived for directed networks only, where the dimensionality reduction has a real computational benefit. In fact, in the case of undirected graphs, it is not possible to remove nodes on the walks from the driver to the targets since information is bound to span the entire network. Similarly, for directed but dense networks, the possibility to focus on specific parts of the network,

and reduce the computational cost, becomes lower regardless of the topology.

To conclude, it is important to mention that extensions of network controllability tools to time-varying frameworks do exist [93, 195]. However, in that case networks would be inferred from gene coexpression and therefore affected by statistical uncertainty due to sample sizes. Further research is needed to seek how to apply network controllability in presence of noisy time-varying connections.

5.1.5 Conclusion

In this study, we introduce a method to quantify the ability of candidate driver nodes to drive the state of a target set within a sparse and directed network. Further, we illustrate how this method works for the molecular network associated with the human macrophage inflammatory response. The obtained results reveal in a principled way the genes that are significantly dysregulated in multiple sclerosis. We hope that this method can contribute to the identification of the key nodes in biological networks to better identify pharmacological targets to counteract human diseases.

5.1.6 Methods

Step-wise target controllability

We introduce a method to identify which target nodes in a network that can be controlled from a single *driver* node. To do so, we start by considering the canonical linear time-invariant dynamics on a directed network described by the adjacency matrix $A \in \mathbb{R}^{N\times N}$

$$\dot{\mathbf{x}}(n) = A\mathbf{x}(n) + B\mathbf{u}(n), \qquad \mathbf{y}(n) = C\mathbf{x}(n) \tag{5.2}$$

where $\mathbf{x}(n) \in \mathbb{R}^N$ describes the state of each node at time n, $B \in \mathbb{R}^N$ specifies the driver node, $\mathbf{u}(n) \in \mathbb{R}^N$ is its external input (or control) signal, $\mathbf{y}(n) \in \mathbb{R}^S$ is the output vector, and $C \in \mathbb{R}^{S\times N}$ is the output matrix identifying the target nodes.

Such a system is controllable if it can be guided from any initial state to any desired final state in finite time, with a suitable choice of input. A necessary and sufficient condition to assess the controllability of **Equation 5.2**, is that the controllability matrix Q

$$Q = \begin{bmatrix} B & AB & A^2B & \cdots & A^{N-1}B \end{bmatrix} \tag{5.3}$$

has full row rank, i.e. $\text{rank}(Q) = N$. That is the Kalman rank condition, which basically verifies the existence of linearly independent rows in Q [82, 144]. If so,

the driver node can reach and control the dynamics of all the other nodes through independent walks of length $N-1$ at maximum.

If it is of interest to control only a target set $\mathcal{T}$ of the network, specified in C and consisting of $S \leq N$ nodes, then **Equation 5.2** can be reduced into a target controllability matrix $Q_{\mathcal{T}} = CQ$ (**Equation 5.1**), where C filters the rows of interest corresponding to the targets. Now, the rank of $Q_{\mathcal{T}}$ gives the number $\tau \leq S$ of nodes in the target set that can be controlled by the driver.

To identify a driver-target configuration, we further introduce a hierarchy among the target nodes, so that we can order and relabel them from the most important one to the least, i.e. $t_1 \succ t_2 \succ ... \succ t_S$. Then we perform the following step-wise procedure for each candidate driver node

- Step 1. *Initialization*
  - Create a temporary empty target set $\mathcal{T}' \leftarrow \{\}$
  - Set the number of controllable targets $\tau \leftarrow 0$
- Step 2. *Repeat until termination criteria are met.* For $j \leftarrow 1, ..., S$ do
  - Add the j-th target node to the target set $\mathcal{T}' \leftarrow \mathcal{T}' \cup \{t_j\}$
  - Build the subgraph containing the nodes on walks from the driver to the targets in $\mathcal{T}'$
  - Compute the rank of the target controllability matrix $Q_{\mathcal{T}'}$
  - *If* rank$(Q_{\mathcal{T}'})$ is full then $\tau \leftarrow \tau + 1$ else $\mathcal{T}' \leftarrow \mathcal{T}' \setminus \{t_j\}$
  - $j \leftarrow j + 1$
- Step 3. *Output* τ *and* $\mathcal{T}'$

Eventually, the *step-wise target control centrality* (τ) is the number of controllable targets in $\mathcal{T}$, and the set $\mathcal{T}'$ contains the τ controllable targets with highest ranking.

Construction of the macrophage activation network

We reconstruct the inflammatory molecular network of the human macrophage by integrating information from the macrophage signal transduction map [139, 143]. This map contains a comprehensive, validated, and annotated map of signal transduction pathways of inflammatory processes in macrophages based on the current literature. To extract molecular interactions from this map, we used the Hermit software [118], which implements automatic reasoning based on logical rules. We

specifically used the rules implemented in the molecular network ontology to infer molecular interactions depending on the process they belong [74, 122]. Because we are interested in the inflammation process, we restricted our analysis to a specific subset of 101 genes with known roles in macrophage pro-inflammatory activation, and for which their regulation in response to pro-inflammatory stimuli could be confirmed in our data set. These genes were classified according to their function in the cell: *sensing*, *signaling*, *transcription* and *secreted* (**Table SA1**), as described in databases such as NCBI Gene [2], UniProt [36] and GeneCards [158]. The full network was thus reduced to only include these genes and their interactions. Due to recent studies, we also opted to exclude two edges (from SOCS3 to IFNGR1 and to IFNGR2) to represent the involved pathways [182].

The resulting network contains $N = 101$ nodes and $L = 211$ unweighted directed edges representing either activation or inhibition between genes. The total degree k of each node in the network is computed by summing the number of incoming and outgoing edges:

$$k_i = \sum_{j=1}^{N} A_{ij} + \sum_{j=1}^{N} A_{ji} \tag{5.4}$$

where $A_{ij} = 1$ if there is an edge between the corresponding genes, and 0 otherwise.

Collection of macrophage mRNA expression data

Collection of blood for the study was approved by the French Ethics committee and the French ministry of research (DC-2012-1535 and AC-2012-1536). Written informed consent was obtained from all study participants. All patients fulfilled diagnostic criteria for multiple sclerosis [166], and individuals (multiple sclerosis patients and healthy donors) with any other inflammatory or neurological disorders were excluded from the study. Patients were included in the study only if they were not undergoing treatment.

Blood was sampled from 8 MS patients and 8 healthy controls in acid citrate dextrose tubes. From blood samples, peripheral blood mononuclear cells were isolated using Ficoll Paque Plus (`www.gelifesciences.com`) and centrifugation (2200 rpm, 20 min). Cells were washed in PBS and RPMI +10% FCS. Monocytes were isolated with anti-CD14 microbeads (`www.miltenyibiotec.com`) and plated in 12-well plates (500000 cells/well) in RPMI +10% FCS and granulocyte-macrophage colony-stimulating factor (500 U/ml) to induce differentiation into macrophages. After 72h, media was replaced with fresh media supplemented with granulocyte-macrophage colony-stimulating factor (500 U/ml) to maintain "alert" macrophages or IFNγ (200 U/ml) + upLPS (10 ng/ml) to induce "pro-inflammatory" activa-

tion. Cells were lysed after 24h and RNA was extracted with RNeasy Mini Kit (`www.qiagen.com`).

Transcriptome sequencing cDNA libraries were prepared using a stranded mRNA polyA selection (Truseq stranded mRNA kit, `www.illumina.com`). For each sample, we performed 60 million single-end, 75 base reads on a NextSeq 500 sequencer (`www.illumina.com`). RNA-Seq data analyses were performed by GenoSplice technology (`www.genosplice.com`). Sequencing, data quality, reads repartition (e.g., for potential ribosomal contamination), and insert size estimation are performed using FastQC [4], Picard-Tools (`http://broadinstitute.github.io/picard/`), Samtools [94] and rseqc [177]. Reads were mapped using STARv2.4.0 [48] on the hg19 Human genome assembly. Gene expression regulation study was performed [125]. Briefly, for each gene present in the FAST DB v2018_1 annotations, reads aligning on constitutive regions (that are not prone to alternative splicing) were counted. Based on these read counts, normalization was performed using DESeq2 [107] in R (v.3.2.5) [165].

Network modeling and data analysis

In the modeling framework described by **Equation 5.2**, matrix A corresponds to the molecular network and represents the time-invariant component of the system. The dynamic component is instead represented by the gene activation response in the healthy and diseased condition (**Figure 5.2b**), computed as the ratio in gene expression between the "pro-inflammatory" and "alert" condition. Specifically, $\mathbf{x}(n)$ represents the gene activation. B is a vector identifying the candidate driver. The control signal $\mathbf{u}(n)$ is out of the scope of this work. The output vector $\mathbf{y}(n)$ and the output matrix C identify the target nodes.

We select the genes belonging to the *secreted* molecules class (**Table SA1**) as our target set $\mathcal{T}$. All the nodes in the other classes are then tested separately as potential driver nodes by computing their step-wise target control centrality (τ). To enhance numerical precision, the logarithmic transformation $\log(q+1)$ is applied to the elements of the target controllability matrix $Q_{\mathcal{T}}$ (**Equation 5.1**).

The hierarchy among the target nodes is established by computing the fold change Δ between the corresponding gene activation in the two groups:

$$\Delta = \frac{\mu_{MS}}{\mu_{HC}} \tag{5.5}$$

where μ_{MS} and μ_{HC} are group-averages for MS patients and healthy controls, respectively, of the gene activation. Nodes with higher Δ absolute values are ranked first. Highly positive Δ values indicate a too strong inflammatory response (over-

activation) in the MS patients with respect to the healthy controls. Highly negative Δ values indicate a too weak inflammatory response (under-activation). We define *dysregulated* genes along the controllable driver-target walks as those for which $|\Delta|$ is above the $75th$ percentile.

We perform a robustness analysis to evaluate the stability of the identified driver nodes to potential errors in the molecular network reconstruction. We simulate attacks with increasing intensity, i.e. up to 20% of the nodes or edges in the network. When removing nodes, we consider the following cases: i) random deletion, ii) preferential removal of high-degree nodes, and iii) preferential removal of low-degree nodes. Preferential attacks are performed by selecting nodes with a probability p proportional to their degree k, i.e. $p \propto k$ for high-degree nodes and $p \propto -k$ for low-degree nodes. When perturbing edges, we test: i) random addition, ii) random deletion, and iii) random rewiring. For each case, we simulated 1000 repetitions and we computed the step-wise target control centrality (τ) for the driver nodes identified in the original network. Then, we report the percentage of nodes that cease to be drivers (i.e. $\tau = 0$), that is, the percentage of nodes that are drivers in our analysis, but are no longer able to control any target in the perturbed case.

5.1.7 Acknowledgments

We would like to thank Prof. Albert-Laszlo Barabasi for his helpful comments and suggestions. The research leading to these results has received funding from the French government under management of Agence Nationale de la Recherche as part of the "Investissements d'avenir" program, reference ANR-19-P3IA-0001 (PRAIRIE 3IA Institute) and reference ANR-10-IAIHU-06 (Agence Nationale de la Recherche-10-IA Institut Hospitalo-Universitaire-6), and from the Inria Project Lab Program (project Neuromarkers). The content is solely the responsibility of the authors and does not necessarily represent the official views of any of the funding agencies.

5.2 Aging brain network

5.2.1 Abstract

Identifying the nodes that may be able to impact the state of a network is a crucial question for many complex interconnected systems. In particular for biological systems, it is a relevant question because it is often possible to only act on single driver nodes and desirable to focus on their ability to control a target node-set of the network. In the case of the brain, there has been a growing interest in characterizing the brain areas most relevant for different cognitive processes. Motivated by the need to understand what may be the single brain regions able to affect specific structural systems, we implemented the step-wise target controllability framework in human connectomes of 171 healthy subjects aged from 5 to 85. We used diffusion tensor imaging (DTI) and functional magnetic resonance imaging (fMRI) to derive, respectively, structural and functional brain networks, and we directed the structural connectomes with a biased random walk process. Results showed that the sensorimotor system tends to have higher centrality than the other systems for all target sets. Interestingly, we found that for most brain regions the centrality tends to be negatively correlated with age, suggesting that their driving potential deteriorates with normal aging. Our work is a preliminary step towards a better understanding of the aging process in healthy subjects.

5.2.2 Introduction

Network controllability has recently received a growing interest from the scientific community, in particular for what concerns biological systems [137, 138, 173, 187]. Great advantages could be provided by a successful application of controllability, which could achieve the ability to steer a system from an initial state to a desired final state. In particular, controllability of brain network has been studied for the possibility to better understand how the brain is able to modify its dynamics in order to enable different cognitive tasks, or cognitive control, and to exploit it in medical treatments [65, 119, 162].

This kind of works addressed questions on whether the brain could theoretically be controlled, and which brain regions may be the best-suited to influence functional states. Centrality measures based on controllability theory were developed and applied to brain networks in order to gain insight in its functioning [65, 130]. These studies have practical applications and can inform medical treatments. For instance, non-invasive transcranial magnetic stimulation [16] and invasive deep brain stimulation [181] are used to treat Alzheimer's disease, Parkinson's disease, stroke, and epilepsy [162].

However, the question of how single brain regions can affect not the whole, but parts, or a subsystem, of the brain is yet to be explored.

In this paper, we aim to study from a control-theoretical perspective the role of single regions of interest (ROIs), detailed in **Tables SB1 and SB2**, in controlling specific target sets. Targets are chosen to be structural systems of the connectome. We analyzed frontal, limbic, temporal, sensorimotor, parietal, and occipital systems, **Figure SB1**. We implemented the step-wise target controllability framework [12], that assumes a linear time invariant dynamics and provides a centrality measure based on the Kalman rank condition [82, 108]. The step-wise target control centrality counts the number of target nodes that can be controlled by a single driver. We used a database containing information on both the functional (fMRI) and the structural (DTI) brain networks for 171 subjects, aged from 5 to 85 [21]. To study more efficiently the controllability property, we directed the connectomes through a biased random walk process [61], **Methods**. We asked how is the centrality measure distributed in the structural systems at the varying of the target set, and how does it change with age. Moreover, we studied what happens in the case of a stroke, or a lesion localized to one hemisphere of the brain.

We found that, independently on the target set, the sensorimotor system presents the highest centrality values, indicating that it may be well suited to control target areas of the brain. On the other hand, the limbic system always presents low centrality values. Interestingly, we found that most brain regions tend to lose importance as time passes. In the case of a lesion, we clearly see that the hemisphere in which the lesion is located loses the most in centrality values. In particular, when the sensorimotor system is attacked global losses are the most severe, instead when damaging the limbic system, global losses in the centrality values are contained.

Our work may be seen as a step forward in the identification of the brain areas that could be most relevant in the aging process, viewed in the original perspective of target controllability.

5.2.3 Results

The sensorimotor system has the largest values of step-wise target control centrality, independently on the target set

We first studied how the step-wise target control centrality (τ) values are distributed across brain systems.

Given the directed brain networks (**Methods**), we analyzed the ability of a single node in controlling a specific target set, **Figure SB1**. To find how many target nodes can be controlled by a single driver, we computed the τ values averaged

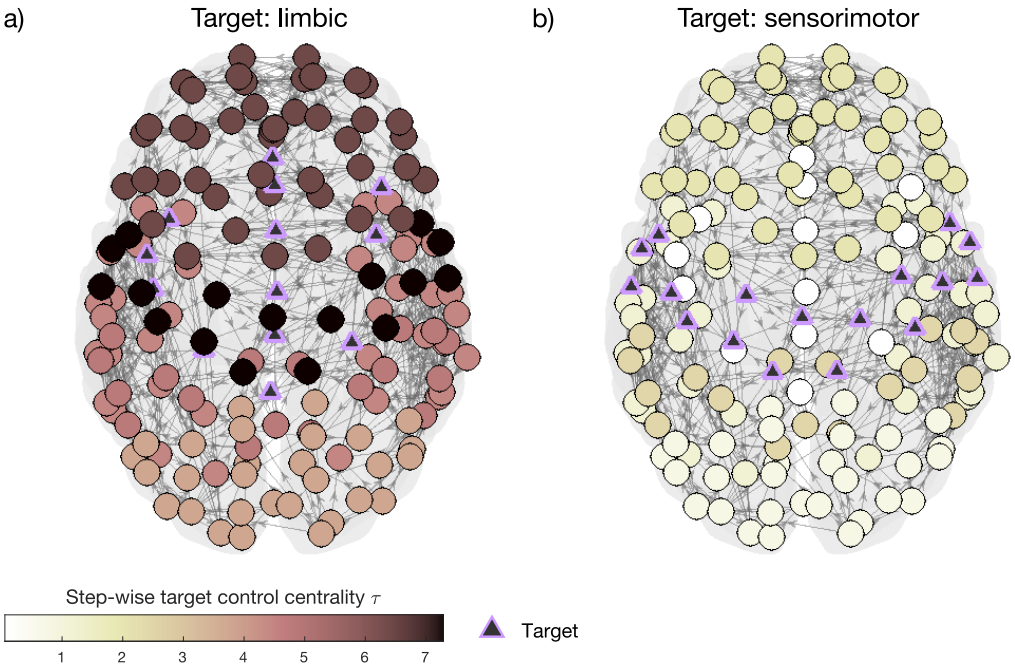


Figure 5.7: **Step-wise target control centrality (τ) averaged across systems and subjects, for the limbic and sensorimotor target sets.** Panel **a)** shows τ centrality values for each brain region when the limbic system is the target set. This case is also exemplary of the case in which one among the frontal, temporal, parietal, or occipital system is the target set. In these configurations, the sensorimotor system has the largest values of centrality, suggesting that it may be well-suited to control the other systems. On the other hand, for all target configurations the limbic system presents low centrality values. Panel **b)** shows τ centrality values for each brain region when the sensorimotor system is the target set. In this case the values are extremely low for all structural systems, suggesting that the sensorimotor system may be difficult to control.

across systems and across subjects in the case of different target systems. Results for all target configurations are reported in **Figure SB2**. We performed a one-way anova [69] to test for the effect of the system on the centrality values, and we found a significant effect (F-values in **Table SB3**).

Two exemplary cases of the distribution of step-wise target control centrality are shown in **Figure 5.7**. We found that the sensorimotor system has the largest centrality values for all target sets, indicating the ability of single sensorimotor brain regions to control large parts of the other systems, see for instance **Figure 5.7a**. Moreover, when the sensorimotor system is the target, all other systems present extremely low τ values, that is, only a small number of nodes in the sensorimotor system can be controlled at the same time by a single brain region in the other structural systems, **Figure 5.7b**. On the other hand, the limbic system presents low centrality values for all target configurations, hinting at its inability to control other systems.

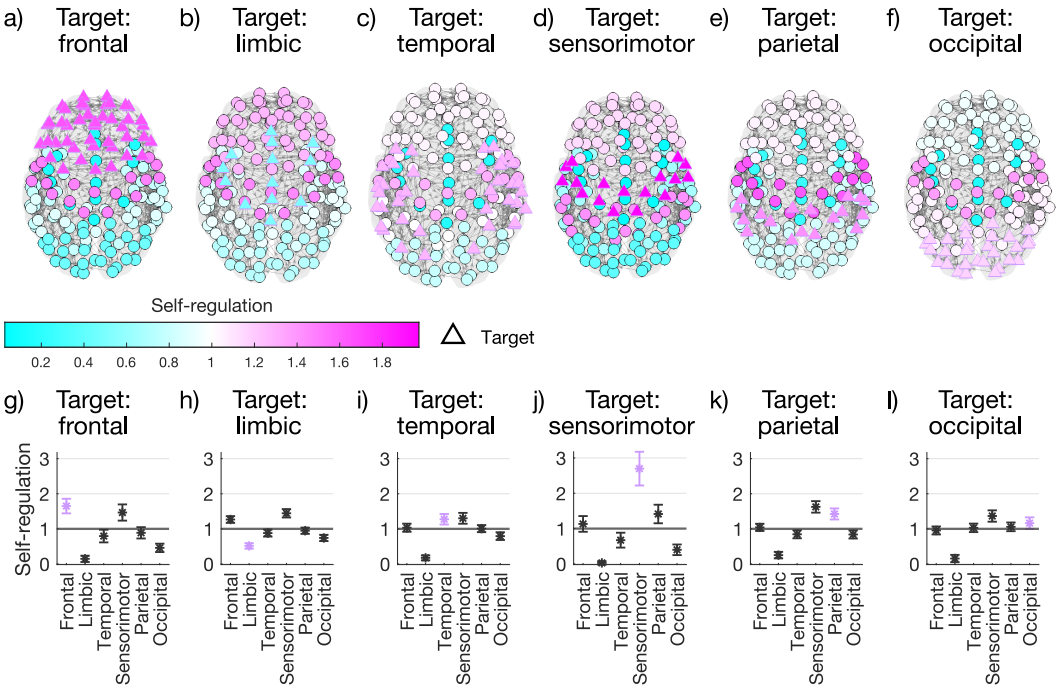


Figure 5.8: **Self-regulation score averaged across systems and subjects, at the varying of the target set.** Self-regulation for a system is computed as the ratio between average τ in that system and total average τ. Thus, self-regulation above (below) 1 indicates that nodes in that system can control more (less) targets than an average node. Across all configurations apart for the limbic system, the target system itself has the largest values of self-regulation, partially explained by the fact that each target node can at least control itself. In all configurations, the limbic system has low self-regulation. Panels **a-f)** show how self-regulation scores are distributed on the connectomes at the varying of the target set. Panels **g-l)** show numerical values of the self-regulation score for the corresponding target sets. Error bars stand for standard error across subjects. The light purple marker denotes the target system.

These results indicate that the sensorimotor system may be well suited control other systems, and at the same time it may be difficult to control.

Self-regulation An additional step in analyzing the power of brain regions in controlling a specific target set is to quantify how much of the centrality is concentrated in that system. It may be of particular interest to study how much can a certain system control itself, that is to understand how much a brain region in the target set can control other ROIs in the same system.

We asked whether the centrality values of the target system are higher than those of the other systems. To quantify it, we computed a self-regulation score, made of the ratio between average τ in a system and average τ in all the network, see **Methods**. When this score is below (above) one, it indicates that, on average, brain regions

in the system under consideration can control a smaller (larger) number of target nodes than an average ROI. We computed self-regulation scores for the six target systems, results are shown in **Figure 5.8**. Then, we performed a one-way anova [69] to test for the effect of the system on the self-regulation scores, and we found a significant effect (F-values in **Table SB3**).

We found that in general the target system has a larger self-regulation score than the other systems, meaning that a node in the target system can, on average, control more targets than an average node. This may be partially explained by the fact that, by construction, nodes in the target system can at least control themselves. An exception to this is given by the limbic system, which always has self-regulation score below one, even when it is the target set. The sensorimotor system consistently presents high values of the self-regulation score, having always values above one.

The step-wise target control centrality of the brain regions deteriorates with age

Next, we asked whether there is a change in the step-wise target control centrality (τ) while growing old. In particular, we were interested in studying which ROIs present a significant change in step-wise target control centrality as age increases.

We computed the Spearman partial correlation between τ and age for each target system, results are reported in **Figure SB3** and **Tables SB4 to SB7**. We corrected for outdegree in order to disentangle the contribution of outgoing connections from that of the centrality. The outdegree is relevant to the controllability property, because outgoing edges make possible to control other nodes. However, because of the procedure we implemented to direct the brain networks, indegree is almost perfectly negatively correlated with outdegree (Spearman rho $= -0.97$, $p < 10^{-6}$), see **Methods**.

When different target sets are considered, the centrality values change. Thus, the correlation may vary for different target sets. ROIs that are significant ($p < 0.05$) in at least one target configuration are shown in **Figure 5.9a**. We focused on those regions that are most consistent across target sets, that is, the regions which consistently present a negative (or positive) correlation across target configurations.

We found that 109 out of 158 regions, 69%, are significant for some target set. Among those, 82 (75%) present a negative trend, and 24 (22%) a positive one. Only for three regions, 3%, (LLG2, RFP10, RPOC1) the correlation changes in sign when different target configurations are considered.

Six regions are consistent across all six target sets, and present a negative correlation. Among those, the *right lateral occipital inferior 3* (RLOCid3) region is in the occipital system, and is known to be part of the dorsal attention network

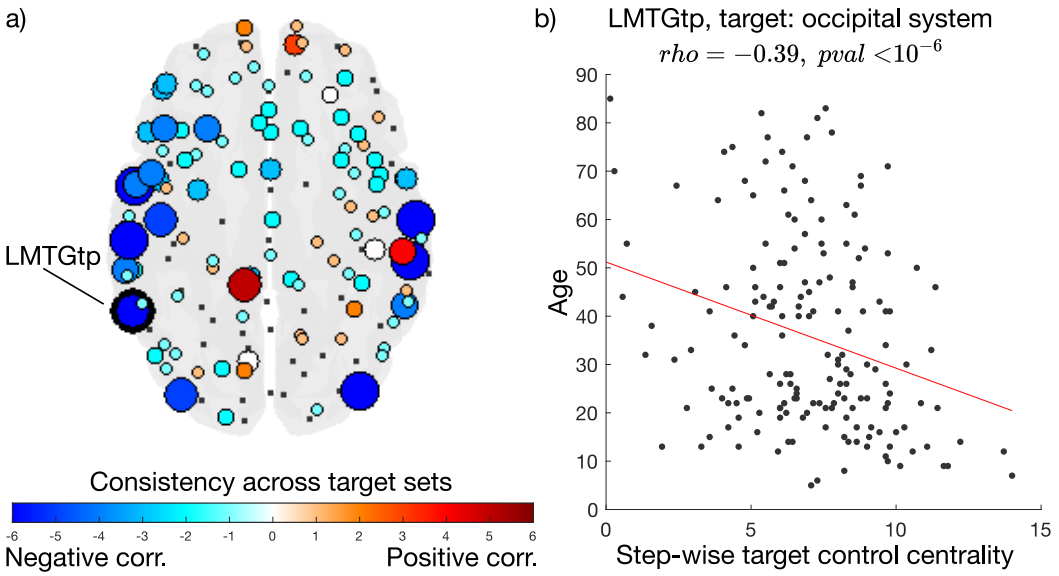


Figure 5.9: **Partial correlation between step-wise target control centrality (τ) and age, corrected for outdegree.** The correction is aimed at disentangling the contribution of outgoing connections and that of the centrality. The outdegree is relevant to the controllability property, in fact outgoing edges make possible to control other nodes. Panel **a)** shows brain regions that present a significant correlation ($p < 0.05$) for at least one target configuration. Bigger and darker dots correspond to ROIs that are consistently significant across multiple target sets. Cool (warm) colors stand for negative (positive) correlations, white identifies ROIs for which the trend changes for different target sets. The ROI with a thick black border is the *left middle temporal temporooccipital* (LMTGtp) region. This region is located in the temporal system, and it is known to be involved in the default mode network. It is the region presenting the strongest negative correlation among the ones consistent for all six target systems. Panel **b)** shows the scatter plot of τ and age for LMTGtp when the target is the occipital system.

[193]. The other regions are in the temporal system (LMTGad, LPT1, LMTGtp, RMTGpd, RITGpd2), and three of them (LMTGad, LMTGtp, RMTGpd) belong to the default mode network (DMN) [43, 186]. Among those, the *left middle temporal temporooccipital* (LMTGtp) region has the strongest negative correlation, the corresponding scatter plot of the τ and age is reported in **Figure 5.9b**.

These results suggest that most brain regions, and in particular those located in the temporal system, tend to decline with age.

Stroke mostly impacts step-wise target control centrality in the damaged hemisphere

We then ask to what extent the centrality values change in the event of a brain attack, or stroke, to the network. To answer this question, we considered the same brain network dataset and we simulated lesions by damaging part of the structural

systems localized in each hemisphere In particular, in brain networks we removed nodes and their associated edges corresponding to the lesion. This led us to consider twelve lesions, left and right sides for the six structural systems, see **Figure SB1**, **Methods**. We computed for each node the step-wise target control centrality (τ) values, and averaged them across systems in the left and right hemisphere, and across subjects. Then, for each ROI we computed the absolute difference between step-wise target control centrality (τ) in the stroke-simulated case, and original τ, shown in **Figure 5.10**.

We found that, in general, systems in the same hemisphere as the lesion suffer a larger loss in step-wise target control centrality, **5.10m-x**. In particular, when attacking any system but the sensorimotor, there were considerable losses in the centrality values, **5.10a-c,e-i,k-l**. In those configurations, the sensorimotor system itself, which started from high τ values in the original case, underwent the most severe losses. Instead, when attacking the sensorimotor, other systems did not sustain considerable losses in the centrality values, **5.10d,j**. Globally the sensorimotor system was the most affected by the stroke simulation and the limbic system was the least affected.

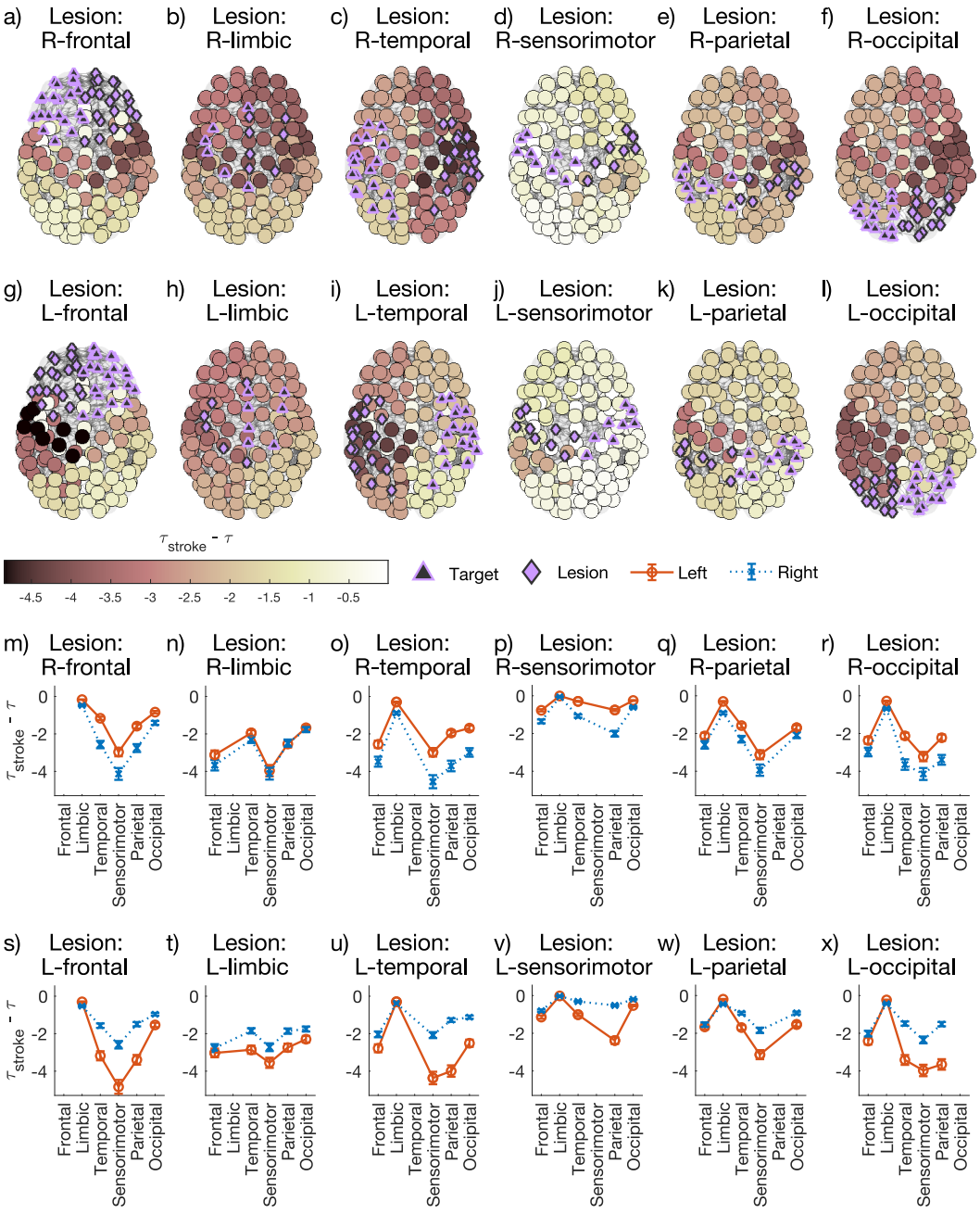


Figure 5.10: **Absolute difference in step-wise target control centrality (τ) between the simulation of a lesion in one hemisphere of the brain and the normal case.** When simulating stroke, the hemisphere most impacted is the one with the lesion. The sensorimotor system appears to be the most affected, while the limbic system is the least affected. Attacking the sensorimotor system led to minor global losses, while attacking any of the other systems led to considerable global losses. Panels **a-l)** show how the absolute difference in centrality values is distributed on the connectomes at the varying of the target set, when the lesion is on the right (panels a-f), or on the left (panels g-l). Panels **m-x)** show averaged values of the absolute difference for the left (solid orange circles) and right (dotted blue crosses) hemispheres for the corresponding target set configurations, when the lesion is on the right (panels m-r), or on the left (panels s-x). Error bars stand for the standard error across subjects.

5.2.4 Discussion

Distribution of step-wise target control centrality in brain networks with different target sets

Controllability refers to the ability of a dynamical system to move from any initial state to any final state in finite time, when applying suitable inputs to well chosen nodes, the drivers [82, 156]. However, in some cases it is neither feasible, nor required to steer the state of the whole system, being more appropriate to exert control on only a subsystem, a subset of the nodes in the network called target set [35, 40, 58]. This may happen, for instance, when developing drug cocktails for multiple target purposes [184, 189].

It is also known that, despite the minimum number of inputs required to achieve control over a system is fixed, the actual composition of the driving set may vary [32, 81]. Even in the case of simple graphs, there may be a huge number of a priori equivalent configurations [75, 101, 174, 192]. Moreover, often in the case of biological networks it is possible to only act on a single node at a time, because of physiological or technical constraints. Thus, the problem of understanding the role played by a single node in driving a part of a system is a relevant question.

In the case of the brain, we may be interested in some subsystems pertaining to specific functions. In this paper, we tested the effect of controlling six different structural systems (frontal, limbic, temporal, sensorimotor, parietal, occipital), denoted as target sets, on the step-wise target control centrality (τ) [12]. This centrality measure is based on the Kalman rank condition for controllability, and counts the number of target nodes that can be simultaneously controlled by a single driver.

We found the sensorimotor system to be deeply implicated, having the highest values of τ, independently on the target set. Since the sensorimotor is a primary system, involved in the default mode network (DMN), we could speculate saying that primary systems are well-suited to implement control of specific tasks. For instance, in the case of vision, the primary visual cortex can activate extended visual pathways and prompt visual perception and scene understanding [119].

Decline of step-wise target control centrality when aging

It is well-known that age is associated with cognitive decline [147]. There is a consensus among the scientific community in identifying a link between aging and a hypoactivation of the prefrontal network, functional disruptions in precuneus, retrosplenial and posterior cingulate cortices, and a compensatory mechanism involving cortical recruitment [167].

Age is also among the primary risk factors for neurodegenerative disorders [147,

190]. For instance, it was shown that Alzheimer's disease patients present a deposition of amyloid plaques in precuneus, retrosplenial and posterior cingulate cortices, which were also found in healthy older adults [127, 141].

The study of the brain as a network and the analysis of topological centrality measures has recently brought insight to the characterization of brain diseases [157]. However, in some cases results have been inconsistent and modality-dependent [41, 67, 188]. Thus the question is still open to find a detailed and specific characterization of the aging process.

In this study, we combined information from both functional (fMRI) and structural (DTI) neuroimaging data, and we analyzed the effect of aging on a measure of centrality based on network controllability. We found regions in the temporal system, and in particular some regions in the DMN (LMTGad, LMTGtp, RMTGpd), and one in the dorsal attention network (RLOCid3), to be deeply affected by aging, in accordance with previous literature studying both healthy subjects [167] and Alzheimer's disease patients [67].

In particular, the sensitivity of DMN to aging effects was studied as disruption of cortical networks caused by amyloid deposition [127], potential gliosis [142], reduced resting-state activity [42], and impaired performance in working memory [148]. The age-related decline in dopamine neurotransmission may contribute to deterioration of functioning of the dorsal attention network, which makes possible sustained attention, one of the most affected cognitive functions that occurs with aging [56].

The relation between brain controllability and aging is currently under-explored, yet it may bring new insight to the question. Since previous studies have addressed a different behavior for long-range and short-range functional hubs [167], it may be of interest to study the effect that target sets at a long or short distance from the driver node have on its controllability properties, taking into account the fact that in order to be controlled simultaneously the targets must be on paths at different lengths from the driver [104].

Effect of lesions on the step-wise target control centrality

In the event of a disease as stroke, or of a lesion in younger ages, it is known that the brain is forced to react to the disruption with a functional reorganization [27, 198].

We studied the effect of a localized attack to one hemisphere of the brain on the step-wise target control centrality. We found that the hemisphere in which occurs the lesion underwent larger losses in the centrality. This may be explained by the fact that the networks we are considering have much more connections within hemispheres than between hemispheres, reflecting the modular structure of the brain [57], see **Figure SB4**. Thus, removing nodes in the lesion when simulating stroke

may lead to segregation of the two hemispheres and to major losses to the attacked side.

We also confirmed that the least amount of losses in centrality values are found when damaging the sensorimotor system. This is consistent with the fact that this is a primary system, densely connected [123, 140, 154], and presenting a high structural redundancy [112]. In future developments, it may be of interest to study how do the losses associated to a lesion vary with the age of the subjects, to test whether our method is consistent with models of brain connectivity.

Methodological considerations

In this work, we exploited the step-wise target controllability framework [12], which relies on a linear time invariant (LTI) dynamics and is based on the Kalman rank condition. While it is known that the brain presents a nonlinear dynamics, the study of linear models has proved to be beneficial in improving our understanding [65, 102, 162]. Specifically, the controllability of a linearized model can inform on the controllability of the nonlinear model [153].

In order to study the controllability of brain networks efficiently, we gave a directionality to the connectomes. Despite the fact that it is not currently possible for neuroimaging techniques to discern the directionality of bundles of axons, it is known that each neuron propagates signals through a well-defined direction from the soma to the end of the axon. Previous efforts to direct a connectome relied on the hypothesis that, given a set of brain networks, edges present for all subjects are the oldest, and any new edge would be directed from the new node to the existing cluster [86]. However, a limitation of this method is that it is strictly dependent on the initial set of networks, and the procedure would not be easily scalable. Other methods based on local navigation and communicability were devised to infer the directionality of neural signaling, but since they operate best for nodes connected by longer paths, they are not well-suited to perform inference for structurally connected nodes [150].

Our work was instead inspired by diffusion processes [1, 6, 62, 135, 183], and despite the fact that they do not allow for a parallel distribution of the information, our method preserves hierarchical and modular properties of the connectomes, **Figure SB5**.

5.2.5 Conclusions

In this work, we studied a method to quantify the ability of candidate driver nodes to drive the state of a target set in directed brain networks. Further, we studied how

this centrality measure is distributed, and how does it change with age and in the event of a lesion. The obtained results reveal in a principled way the systems that are well-suited to drive specific target sets, and the tendency of most brain regions to lose importance as growing old. We hope that further developments of this method and other network controllability measures could contribute to the identification of the key nodes in biological networks to better identify targets of brain stimulation to counteract human diseases.

5.2.6 Methods

Structural and functional brain networks

We used data from the NKI-Rockland database [126], and selected 171 subjects, aged from 5 to 85 [21]. For each subject we had access to both structural (DTI) and functional (fMRI) data. We performed both data acquisition and preprocessing, and the construction of the brain networks as in Guillon et al. [67].

We directed the structural connectomes in order to study their controllability properties in a more efficient way.

Directing structural brain networks Starting from the DTI network, first we applied a logarithmic transformation ($\log(x+1)$) to make the weights more homogeneous. Then, we performed a biased random walk [61] to direct each edge from the node with lower strength to the node with higher strength. The probability P_{ji} to go from node i to node j, can be computed as

$$P_{ji} = \frac{w_{ji}}{\sum_h w_{ih}} \tag{5.6}$$

where $[w]_{ji}$ is the symmetric, weighted adjacency matrix of the structural network. For each pair of nodes i and j, we had two directed edges with weights P_{ij} and P_{ji}. An example of the resulting matrices is shown in **Figure SB5**. We chose one direction by keeping only the highest probability (we kept both if they were equal). At last, we applied a topological criterion for filtering information [55], to lower the connection density of the networks. We considered for each subject networks with mean degree k varying from 1 to 14.

Indegree and outdegree of the resulting structural systems averaged across mean degree k and across subjects are shown in **Figure SB6**. Notice that, by construction, indegree is almost perfectly negatively correlated with outdegree (Spearman rho $= -0.97$, $p < 10^{-6}$).

Density of the brain networks The subject with the lowest number of connections after applying the procedure to direct the connectomes presented 2350 edges, that is a density of $\frac{E}{N(N-1)} = \frac{2350}{158 \cdot 157} = 0.0947$. Since $\frac{N \cdot k}{N(N-1)} = \frac{15}{157} = 0.0955$, it was not possible to consider mean degrees above 14.

With a mean degree of $k = 14$, the maximum density considered was $\frac{N \cdot k}{N(N-1)} = \frac{14}{157} = 0.0892$, and the minimum density considered was $\frac{N \cdot k}{N(N-1)} = \frac{1}{157} = 0.0064$.

Stroke simulation We simulated stroke by producing a localized lesion in one hemisphere. We attacked brain regions either in the left or right hemisphere of a system, removing the corresponding nodes in the network. To study differences in centrality values between hemispheres, we considered the same six target systems, but damaging one side. For instance, for the frontal target system we computed the centrality values in two cases: when attacking the right hemisphere, obtaining a lesion in the right frontal system and using as target the left frontal system, and when attacking the left hemisphere. This procedure had not only the effect to lower the number of nodes in the target set, but also to change the topology of the network, disrupting connections to and from the damaged areas.

Step-wise target controllability

We implemented the step-wise target controllability framework [12], that analyses the single-input target controllability problem, in which the interest is to study the role of a single driver node in controlling a target set of the system.

We assume the linear time invariant (LTI) dynamics

$$\dot{\mathbf{x}}(n) = A\mathbf{x}(n) + B\mathbf{u}(n), \qquad \mathbf{y}(n) = C\mathbf{x}(n) \tag{5.7}$$

where $\mathbf{x}(n) \in \mathbb{R}^N$ describes the state of each node at time n, A is the adjacency matrix of the network, $B \in \mathbb{R}^N$ specifies the driver node that will receive an external input, $\mathbf{u}(n) \in \mathbb{R}^N$ is its external input (or control) signal, $\mathbf{y}(n) \in \mathbb{R}^S$ is the output vector, and $C \in \mathbb{R}^{S \times N}$ is the output matrix identifying the target nodes.

We say that such a system is controllable if it can be guided from any initial state to any desired final state in finite time, with a suitable choice of input. Consistently, the driver can control the target nodes identified in matrix C if it is possible to guide the state of the target nodes from any initial state to any desired final state in finite time, with a suitable choice of the input injected in the driver node.

In this specific application, the state of the network was represented by the fMRI values, the adjacency matrix corresponded to the directed structural connectome for each subject (averaged across mean degree k from 1 to 14), the target sets were

the structural systems (frontal, limbic, temporal, sensorimotor, parietal, occipital, **Figure SB1**), ROIs in the target system were ranked according to the mean strength across subjects in the functional connectome.

The step-wise target control centrality (τ) computes how many target nodes can be controlled simultaneously by a single driver, and provides the highest-ranked controllable configuration of targets.

Self-regulation score To evaluate the contribution of each structural system to the step-wise target control centrality, and in particular to account for how much nodes in the target system could influence other target ROIs, we devised a self-regulation score. Self-regulation for a system was computed as the ratio between average τ in that system and total average τ. In particular, the self-regulation score for the target system is

$$\frac{\frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}}(\tau_i)}{\frac{1}{N}\sum_{j\in\mathcal{V}}(\tau_j)} \tag{5.8}$$

where $\mathcal{T}$ is the target set, $|\mathcal{T}|$ its cardinality, that is the number of elements in the set, $\mathcal{V}$ is the set of all N ROIs in the brain network, and τ_i is the step-wise target control centrality of each node i.

Self-regulation above (below) 1 indicates that nodes in the considered system can control more (less) targets than an average node.

5.2.7 Acknowledgments

The research leading to these results has received funding from the French government under management of Agence Nationale de la Recherche as part of the "Investissements d'avenir" program, reference ANR-19-P3IA-0001 (PRAIRIE 3IA Institute) and reference ANR-10-IAIHU-06 (Agence Nationale de la Recherche-10-IA Institut Hospitalo-Universitaire-6), and from the Inria Project Lab Program (project Neuromarkers). The content is solely the responsibility of the authors and does not necessarily represent the official views of any of the funding agencies.

Chapter 6

Conclusions

In this thesis, we studied the problem of controlling complex networks from a practical perspective, combining both theoretical and experimental aspects. We proposed an optimized centrality measure based on network controllability to have a more robust estimate of the driver nodes and their "controlling" power.

This led us to uncover relevant functional interactions between nodes in two different biological applications. In the case of macrophage activation in multiple sclerosis, our results indicate the presence of causal relations between driver nodes and their controllable targets, suggesting our method as a possible alternative to traditional dimensionality reduction techniques. In the case of aging brain networks, we were able to identify brain regions previously implicated in degenerative processes. Taken together these results suggest the potential in the application of solutions inspired by controllability theory to biological systems. However, careful attention should be paid to the technical details of the implementation, and more research is needed to generalize these principles.

Recent advances in both network theory and controllability theory inspired questions concerning the fundamental controllability principles of complex networks, and their balance between structure, control and function [9, 11, 18, 162]. These kinds of questions have driven the research for a better understanding of biological systems such as the brain [119, 187], gene regulatory networks [137], and protein-protein interaction networks [138, 173]. Our work can be seen as a step forward in the direction of adopting a control-theoretical perspective to biological applications.

We proposed the step-wise target control centrality to study the role of single driver nodes in influencing specific target sets in a network. We addressed the single-input target control problem, providing an exact method able to efficiently identify a controllable configuration of targets in sparse, directed networks.

Our method relies on the strong hypothesis that the system can be described by

a linear time invariant dynamics. While it is known that linear models can provide useful information [65, 102], this kind of dynamics is not always appropriate for biological systems such as the brain [78]. However, our algorithm is flexible, and future developments may include different criteria to perform the controllability test, such as the Popov-Belevitch-Hautus criterion [24, 156], or the Gramian condition [25]. A different controllability test altogether, for instance a feedback vertex set [191] approach, able to deal with nonlinear dynamics, may be implemented.

A seemingly subjective choice in our algorithm is the choice of an ordering for the target set. Future work may involve the study of the effect of different orderings on standard graph models such as Erdös-Rényi random graphs [54] and Barabási-Albert scale free networks [7], in order to inform the choice in a principled way.

To understand complex biological systems we require expertise from different fields, studying them from a control-theoretical perspective offers the promise to a new insight.

Scientific production

Submitted journal papers

- **Bassignana G.**, Fransson J., Henry V., Colliot O., Zujovic V., De Vico Fallani F., *Step-wise target controllability identifies dysregulated pathways of macrophage networks in multiple sclerosis.* 2020. arXiv:2003.08913v3. [12]

In preparation journal papers

- **Bassignana G.**, Colliot O., De Vico Fallani F., *A step-wise criterion to determine target control centrality in the aging brain network.*

Contribution to schools

- Oral presentation at *Complex networks: from socio-economic systems to biology and brain* workshop, Lipari, Italy, July 2019.

Contribution to conferences

- Poster presentation at *Network neuroscience* online conference and oral presentation at *Network neuroscience* satellite, September 2020.
- Oral presentations at *Network science (NetSci)* conference and *Network neuroscience* satellite, Burlington, VT, May 2019.
- Oral presentation at *IEEE workshop on complexity in engineering*, Florence, Italy, October 2018.
- Poster presentation at *Conference on Complex Systems*, Thessaloniki, Greece, September 2018.
- Lightning talk and poster presentation at *Network science (NetSci)* conference, Paris, France, June 2018.

Conference attendance

- IEEE Conference on Decision and Control, Nice, France, December 2019.

Other scientific production

Published journal papers

- Ansart M., Epelbaum S., **Bassignana G.**, Bône A., Bottani S., Cattai T., Couronne R., Faouzi J., Koval I., Louis M., Thibeau-Sutre E., Wen J., Wild A., Burgos N., Dormont D., Colliot O., and Durrleman S., *Predicting the Progression of Mild Cognitive Impairment Using Machine Learning: A Systematic, Quantitative and Critical Review. Medical Image Analysis.* October 2020. [5]

Submitted journal papers

- Gonzalez-Astudillo J., Cattai T., **Bassignana G.**, Corsi M.-C., and De Vico Fallani F., *Network-based brain computer interfaces: principles and applications.* 2020. arXiv:2006.13187. [63]

In preparation journal papers

- Henry V., **Bassignana G.**, Lejeune F. X., Zujovic V., De Vico Fallani F., Dameron O., Moszer I., Colliot O., *OWL ontology and SWRL rules applied to systems biology networks conciliation.*

Appendix A

Supplementary material for Section 5.1

A.1 Supplementary figures

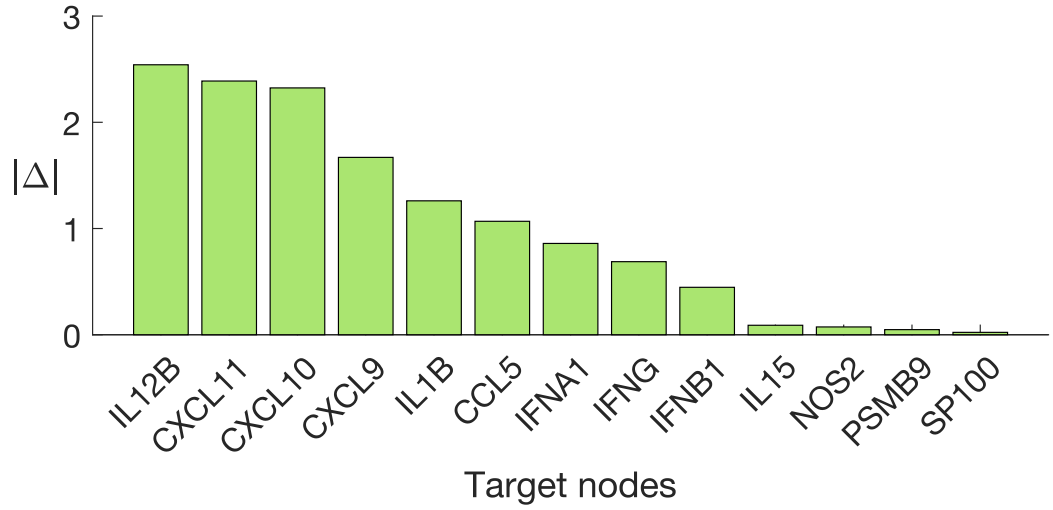


Figure SA1: **Hierarchy among target genes.** Genes corresponding to the 13 secreted molecules are ranked according to the absolute value of fold change Δ in the gene activation between the multiple sclerosis (MS) group and the healthy control (HC) group (**Methods**).

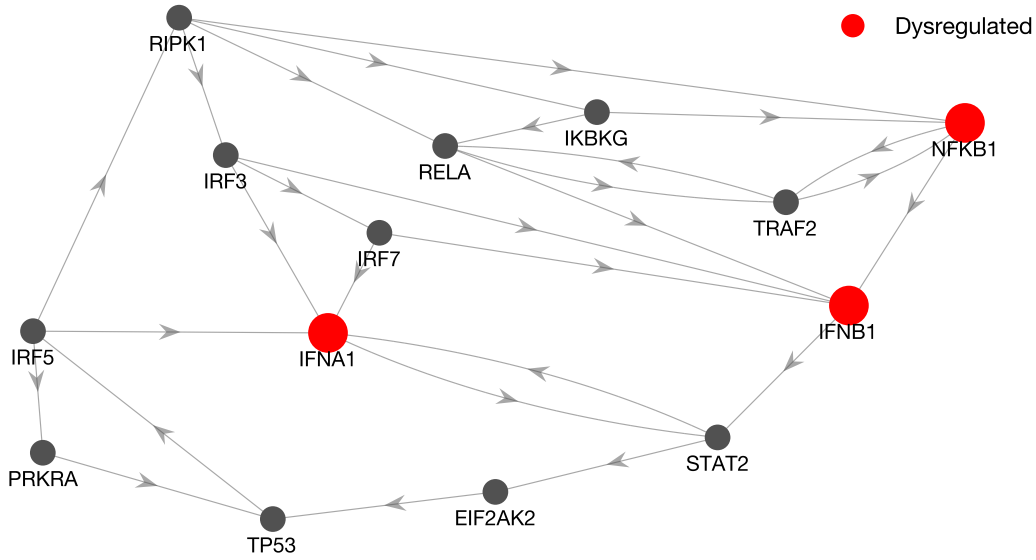


Figure SA2: Subnetwork illustrating the feedback cycle between dysregulated genes IFNA1, IFNB1 and NFKB1. The three nodes belong to the only strongly connected component (a subnetwork in which every node is reachable from any other node) of the network having more than two nodes. It plays, thus, a central role in the network topology.

A.2 Supplementary tables

Gene	Class	Gene (cont.)	Class (cont.)
AIFM1	transcription	IL15	secreted molecule
APAF1	signaling	IL1B	secreted molecule
ATF2	transcription	IRF1	transcription
BAK1	signaling	IRF2	transcription
BAX	signaling	IRF3	transcription
BBC3	transcription	IRF4	transcription
BCL2	signaling	IRF5	transcription
BCL2L1	signaling	IRF7	transcription
BID	signaling	IRF8	transcription
BIRC2	transcription	IRF9	transcription
BIRC3	transcription	JAK1	signaling
CASP1	signaling	JAK2	signaling
CASP10	signaling	MDM2	signaling
CASP2	signaling	MYD88	signaling
CASP3	signaling	NFKB1	transcription
CASP4	signaling	NFKB2	transcription
CASP6	signaling	NFKBIA	transcription
CASP7	signaling	NOS2	secreted molecule
CASP8	signaling	OAS1	signaling
CASP9	signaling	PARP1	signaling
CCL5	secreted molecule	PLSCR1	transcription
CFLAR	transcription	PRKACA	signaling
CHUK	signaling	PRKCD	signaling
CIITA	transcription	PRKRA	signaling
CXCL10	secreted molecule	PSMB9	secreted molecule
CXCL11	secreted molecule	RELA	transcription
CXCL9	secreted molecule	RELB	transcription
CYCS	signaling	RIPK1	signaling
DFFA	signaling	SOCS1	signaling
DIABLO	signaling	SOCS3	signaling
EIF2AK2	signaling	SOD2	signaling
ENDOG	transcription	SP100	secreted molecule
EP300	transcription	SPI1	transcription
FADD	signaling	STAT1	transcription
FAS	sensing	STAT2	transcription
FASLG	signaling	TBK1	signaling
GBP1	signaling	TICAM1	signaling
HLAB	sensing	TICAM2	signaling
ICAM1	sensing	TLR3	sensing
IFNA1	secreted molecule	TLR7	sensing
IFNAR1	sensing	TLR9	sensing
IFNAR2	sensing	TNFRSF10A	sensing
IFNB1	secreted molecule	TNFRSF10B	sensing
IFNG	secreted molecule	TNFSF10	sensing
IFNGR1	sensing	TP53	transcription
IFNGR2	sensing	TRADD	signaling
IKBKB	transcription	TRAF1	signaling
IKBKE	transcription	TRAF2	signaling
IKBKG	transcription	TRAF6	signaling
IL12B	secreted molecule	TYK2	signaling
		XIAP	transcription

Table SA1: List of all node genes and associated class depending on their functional role.

Centrality	Driver	Controllable targets
10	SOCS1	CCL5; CXCL10; CXCL11; IFNA1; IFNB1; CXCL9; IFNG; IL12B; IL1B; NOS2
10	SOCS3	CCL5; CXCL10; CXCL11; IFNA1; IFNB1; CXCL9; IFNG; IL12B; IL1B; NOS2
9	IFNGR1	CCL5; CXCL10; CXCL11; IFNA1; IFNB1; CXCL9; IL12B; IL1B; NOS2
9	IFNGR2	CCL5; CXCL10; CXCL11; IFNA1; IFNB1; CXCL9; IL12B; IL1B; NOS2
9	JAK1	CCL5; CXCL10; CXCL11; IFNA1; IFNB1; CXCL9; IL12B; IL1B; NOS2
9	JAK2	CCL5; CXCL10; CXCL11; IFNA1; IFNB1; CXCL9; IL12B; IL1B; NOS2
9	PRKCD	CCL5; CXCL10; CXCL11; IFNA1; IFNB1; CXCL9; IL12B; IL1B; NOS2
9	STAT1	CCL5; CXCL10; CXCL11; IFNA1; IFNB1; CXCL9; IL12B; IL1B; NOS2
8	IRF1	CCL5; CXCL10; CXCL11; IFNA1; IFNB1; CXCL9; IL12B; IL1B
8	PRKACA	CCL5; CXCL10; CXCL11; IFNA1; IFNB1; CXCL9; IL12B; IL1B
7	IRF2	CCL5; CXCL10; CXCL11; IFNA1; IFNB1; IL1B; NOS2
1	CASP1	IL12B
1	IRF8	IL12B
1	SPI1	IL12B

Table SA2: Driver nodes and their controllable targets. The majority of driver nodes (omitted from the table) are able to control 5 target nodes (CCL5, CXCL10, CXCL11, IFNA1, IFNB1).

Class	Nodes	Drivers	Fraction of drivers in the class
Sensing	13	7	0.54
Signaling	44	20	0.45
Transcription	31	18	0.58

Table SA3: Distribution of driver nodes across different classes of genes.

rho HC	IL12B	CXCL11	CXCL10	CXCL9	IL1B	CCL5	IFNA1	IFNG	IFNB1	NOS2
ATF2	-	0.262	-0.024	-	-	0.762	0.381	-	0.190	-
CASP1	-	-	-	-	0.667	-	-	-	-	-
CHUK	-	0.643	0.429	-	-	0.286	-0.190	-	0.000	-
EIF2AK2	-	0.690	0.738	-	-	0.738	0.452	-	0.286	-
EP300	-	0.524	0.333	-	-	0.619	0.024	-	-0.143	-
FADD	-	0.381	0.190	-	-	0.833	0.286	-	0.024	-
IFNAR1	-	0.452	0.143	-	-	0.714	0.381	-	0.167	-
IFNAR2	-	0.571	0.333	-	-	0.976	0.429	-	0.143	-
IFNGR1	-0.214	-0.262	-0.524	-0.429	0.071	0.000	0.048	-	0.190	0.204
IFNGR2	0.833	0.524	0.238	0.262	0.738	0.905	0.333	-	0.024	0.910
IKBKB	-	-0.048	-0.119	-	-	0.524	0.262	-	-0.119	-
IKBKE	-	0.405	0.310	-	-	0.738	0.452	-	0.095	-
IKBKG	-	0.286	0.024	-	-	0.595	0.190	-	0.214	-
IRF1	0.810	0.595	0.381	0.500	0.905	0.952	0.595	-	0.405	-
IRF2	-	0.357	0.119	-	0.738	0.905	0.595	-	0.286	0.838
IRF3	-	-0.167	0.167	-	-	-0.595	-0.286	-	-0.071	-
IRF5	-	-0.310	-0.190	-	-	0.095	0.119	-	-0.357	-
IRF7	-	0.929	0.929	-	-	0.643	0.167	-	0.286	-
IRF8	-	-	-	-	0.357	-	-	-	-	-
IRF9	-	0.476	0.595	-	-	0.500	0.357	-	0.190	-
JAK1	0.476	-0.024	-0.262	-0.143	0.381	0.571	0.167	-	-0.167	0.431
JAK2	0.595	0.810	0.667	0.548	0.738	0.690	0.333	-	0.310	0.778
MYD88	-	-0.095	0.143	-	-	-0.429	0.286	-	0.476	-
NFKB1	-	0.286	0.071	-	-	0.786	0.310	-	0.048	-
PRKACA	-0.833	-0.476	-0.357	-0.190	-0.548	-0.714	-0.024	-	0.310	-
PRKCD	0.143	-0.119	-0.405	-0.190	0.286	0.381	0.190	-	0.024	0.431
PRKRA	-	-0.548	-0.381	-	-	-0.810	-0.571	-	-0.500	-
RELA	-	0.190	-0.071	-	-	0.786	0.405	-	0.119	-
RIPK1	-	0.333	0.167	-	-	0.881	0.643	-	0.333	-
SOCS1	0.690	0.500	0.238	0.381	0.714	0.857	0.452	0.228	0.167	0.898
SOCS3	0.357	0.119	-0.143	0.238	0.619	0.571	0.214	0.736	0.167	0.323
SPI1	-	-	-	-	-0.452	-	-	-	-	-
STAT1	0.214	0.643	0.762	0.762	0.310	0.190	0.095	-	0.190	0.156
STAT2	-	0.738	0.905	-	-	0.357	0.310	-	0.381	-
TBK1	-	0.476	0.262	-	-	0.833	0.595	-	0.476	-
TICAM1	-	0.143	0.238	-	-	0.429	0.143	-	-0.214	-
TICAM2	-	0.571	0.310	-	-	0.810	0.286	-	0.214	-
TLR3	-	0.476	0.643	-	-	0.452	0.500	-	0.286	-
TLR7	-	0.286	0.595	-	-	-0.262	0.048	-	0.214	-
TLR9	-	-0.310	-0.190	-	-	-0.690	-0.452	-	-0.381	-
TP53	-	-0.024	0.262	-	-	-0.476	-0.071	-	0.000	-
TRADD	-	0.190	-0.048	-	-	0.643	0.000	-	-0.310	-
TRAF2	-	0.667	0.381	-	-	0.833	-0.024	-	-0.190	-
TRAF6	-	0.214	-0.095	-	-	0.714	0.119	-	-0.167	-
TYK2	-	-0.143	0.095	-	-	-0.786	-0.595	-	-0.238	-

Table SA4: Spearman cross-correlation values for healthy controls (HC) for controllable driver-target node pairs. Values for pairs that were not controllable are not reported (-). Targets IL15, PSMB9, and SP100 could not be controlled by any driver and are omitted from the table.

p values HC	IL12B	CXCL11	CXCL10	CXCL9	IL1B	CCL5	IFNA1	IFNG	IFNB1	NOS2
ATF2	-	0.536	0.977	-	-	0.037	0.360	-	0.665	-
CASP1	-	-	-	-	0.083	-	-	-	-	-
CHUK	-	0.096	0.299	-	-	0.501	0.665	-	1.000	-
EIF2AK2	-	0.069	0.046	-	-	0.046	0.267	-	0.501	-
EP300	-	0.197	0.428	-	-	0.115	0.977	-	0.752	-
FADD	-	0.360	0.665	-	-	0.015	0.501	-	0.977	-
IFNAR1	-	0.267	0.752	-	-	0.058	0.360	-	0.703	-
IFNAR2	-	0.151	0.428	-	-	0.000	0.299	-	0.752	-
IFNGR1	0.619	0.536	0.197	0.299	0.882	1.000	0.935	-	0.665	0.629
IFNGR2	0.015	0.197	0.582	0.536	0.046	0.005	0.428	-	0.977	0.003
IKBKB	-	0.935	0.793	-	-	0.197	0.536	-	0.793	-
IKBKE	-	0.327	0.462	-	-	0.046	0.267	-	0.840	-
IKBKG	-	0.501	0.977	-	-	0.132	0.665	-	0.619	-
IRF1	0.022	0.132	0.360	0.216	0.005	0.001	0.132	-	0.327	-
IRF2	-	0.389	0.793	-	0.046	0.005	0.132	-	0.501	0.013
IRF3	-	0.703	0.703	-	-	0.132	0.501	-	0.882	-
IRF5	-	0.462	0.665	-	-	0.840	0.793	-	0.389	-
IRF7	-	0.002	0.002	-	-	0.096	0.703	-	0.501	-
IRF8	-	-	-	-	0.389	-	-	-	-	-
IRF9	-	0.243	0.132	-	-	0.216	0.389	-	0.665	-
JAK1	0.243	0.977	0.536	0.752	0.360	0.151	0.703	-	0.703	0.285
JAK2	0.132	0.022	0.083	0.171	0.046	0.069	0.428	-	0.462	0.030
MYD88	-	0.840	0.752	-	-	0.299	0.501	-	0.243	-
NFKB1	-	0.501	0.882	-	-	0.028	0.462	-	0.935	-
PRKACA	0.015	0.243	0.389	0.665	0.171	0.058	0.977	-	0.462	-
PRKCD	0.752	0.793	0.327	0.665	0.501	0.360	0.665	-	0.977	0.285
PRKRA	-	0.171	0.360	-	-	0.022	0.151	-	0.216	-
RELA	-	0.665	0.882	-	-	0.028	0.327	-	0.793	-
RIPK1	-	0.428	0.703	-	-	0.007	0.096	-	0.428	-
SOCS1	0.069	0.216	0.582	0.360	0.058	0.011	0.267	0.583	0.703	0.005
SOCS3	0.389	0.793	0.752	0.582	0.115	0.151	0.619	0.050	0.703	0.432
SPI1	-	-	-	-	0.267	-	-	-	-	-
STAT1	0.619	0.096	0.037	0.037	0.462	0.665	0.840	-	0.665	0.716
STAT2	-	0.046	0.005	-	-	0.389	0.462	-	0.360	-
TBK1	-	0.243	0.536	-	-	0.015	0.132	-	0.243	-
TICAM1	-	0.752	0.582	-	-	0.299	0.752	-	0.619	-
TICAM2	-	0.151	0.462	-	-	0.022	0.501	-	0.619	-
TLR3	-	0.243	0.096	-	-	0.267	0.216	-	0.501	-
TLR7	-	0.501	0.132	-	-	0.536	0.935	-	0.619	-
TLR9	-	0.462	0.665	-	-	0.069	0.267	-	0.360	-
TP53	-	0.977	0.536	-	-	0.243	0.882	-	1.000	-
TRADD	-	0.665	0.935	-	-	0.096	1.000	-	0.462	-
TRAF2	-	0.083	0.360	-	-	0.015	0.977	-	0.665	-
TRAF6	-	0.619	0.840	-	-	0.058	0.793	-	0.703	-
TYK2	-	0.752	0.840	-	-	0.028	0.132	-	0.582	-

Table SA5: Spearman cross-correlation p-values for healthy controls (HC) for controllable driver-target node pairs. Values for pairs that were not controllable are not reported (-). Targets IL15, PSMB9, and SP100 could not be controlled by any driver and are omitted from the table.

rho MS	IL12B	CXCL11	CXCL10	CXCL9	IL1B	CCL5	IFNA1	IFNG	IFNB1	NOS2
ATF2	-	-0.571	-0.833	-	-	-0.143	0.708	-	0.293	-
CASP1	-	-	-	-	-0.024	-	-	-	-	-
CHUK	-	0.643	0.476	-	-	0.452	-0.244	-	-0.317	-
EIF2AK2	-	-0.333	-0.214	-	-	-0.667	0.464	-	0.488	-
EP300	-	-0.548	-0.119	-	-	-0.452	0.073	-	0.073	-
FADD	-	0.143	-0.286	-	-	0.619	-0.024	-	0.122	-
IFNAR1	-	0.262	-0.048	-	-	0.548	0.000	-	-0.268	-
IFNAR2	-	-0.429	-0.286	-	-	-0.429	0.195	-	-0.024	-
IFNGR1	0.095	-0.167	0.167	-0.333	0.310	0.095	-0.024	-	0.024	0.190
IFNGR2	-0.310	-0.071	-0.619	0.095	0.048	0.595	0.049	-	-0.122	0.238
IKBKB	-	-0.476	-0.381	-	-	-0.643	0.586	-	0.439	-
IKBKE	-	0.476	-0.048	-	-	0.929	-0.293	-	-0.146	-
IKBKG	-	0.190	0.048	-	-	0.048	-0.244	-	-0.195	-
IRF1	0.381	-0.357	-0.595	-0.595	0.524	0.167	0.708	-	0.195	-
IRF2	-	-0.643	-0.952	-	-0.024	-0.071	0.464	-	0.171	0.452
IRF3	-	0.452	0.786	-	-	-0.024	-0.683	-	-0.220	-
IRF5	-	-0.524	-0.452	-	-	-0.500	0.342	-	0.122	-
IRF7	-	0.500	0.571	-	-	-0.095	-0.171	-	0.293	-
IRF8	-	-	-	-	0.476	-	-	-	-	-
IRF9	-	-0.405	-0.143	-	-	-0.762	0.220	-	0.220	-
JAK1	0.595	-0.238	-0.452	-0.548	0.571	0.048	0.756	-	0.366	0.786
JAK2	0.190	0.476	0.500	0.357	-0.167	0.238	-0.586	-	0.024	-0.429
MYD88	-	-0.762	-0.333	-	-	-0.857	0.171	-	-0.244	-
NFKB1	-	0.071	-0.405	-	-	0.405	0.732	-	0.683	-
PRKACA	-0.095	0.500	0.810	0.381	0.119	0.214	-0.512	-	-0.171	-
PRKCD	-0.190	-0.095	-0.381	-0.405	0.595	0.571	0.171	-	-0.073	0.500
PRKRA	-	0.119	0.571	-	-	-0.381	-0.073	-	0.293	-
RELA	-	-0.190	-0.714	-	-	0.143	0.683	-	0.464	-
RIPK1	-	-0.738	-0.714	-	-	-0.524	0.659	-	0.317	-
SOCS1	0.333	-0.143	-0.190	-0.405	0.143	-0.167	-0.073	-0.275	0.659	0.119
SOCS3	0.333	-0.048	-0.524	-0.310	0.833	0.429	0.756	0.503	0.366	0.929
SPI1	-	-	-	-	-0.262	-	-	-	-	-
STAT1	-0.119	0.571	0.833	0.619	-0.286	0.143	-0.708	-	-0.293	-0.667
STAT2	-	-0.119	0.190	-	-	-0.714	0.220	-	0.586	-
TBK1	-	-0.571	-0.714	-	-	-0.429	0.732	-	0.415	-
TICAM1	-	-0.143	-0.595	-	-	0.119	0.464	-	0.634	-
TICAM2	-	-0.095	0.048	-	-	-0.286	0.634	-	0.634	-
TLR3	-	-0.690	-0.286	-	-	-0.286	-0.415	-	-0.488	-
TLR7	-	-0.286	0.190	-	-	-0.905	0.268	-	0.049	-
TLR9	-	0.524	0.476	-	-	0.619	-0.586	-	-0.512	-
TP53	-	0.643	0.714	-	-	0.024	-0.634	-	-0.122	-
TRADD	-	-0.476	-0.548	-	-	-0.167	0.293	-	-0.122	-
TRAF2	-	-0.071	-0.619	-	-	0.595	0.049	-	-0.122	-
TRAF6	-	-0.429	-0.690	-	-	-0.286	0.756	-	0.293	-
TYK2	-	0.238	0.381	-	-	-0.333	-0.293	-	0.073	-

Table SA6: Spearman cross-correlation values for multiple sclerosis (MS) patients for controllable driver-target node pairs. Values for pairs that were not controllable are not reported (-). Targets IL15, PSMB9, and SP100 could not be controlled by any driver and are omitted from the table.

p values MS	IL12B	CXCL11	CXCL10	CXCL9	IL1B	CCL5	IFNA1	IFNG	IFNB1	NOS2
ATF2	-	0.151	0.015	-	-	0.752	0.059	-	0.490	-
CASP1	-	-	-	-	0.977	-	-	-	-	-
CHUK	-	0.096	0.243	-	-	0.267	0.567	-	0.455	-
EIF2AK2	-	0.428	0.619	-	-	0.083	0.260	-	0.227	-
EP300	-	0.171	0.793	-	-	0.267	0.887	-	0.889	-
FADD	-	0.752	0.501	-	-	0.115	0.979	-	0.795	-
IFNAR1	-	0.536	0.935	-	-	0.171	1.000	-	0.531	-
IFNAR2	-	0.299	0.501	-	-	0.299	0.650	-	0.977	-
IFNGR1	0.840	0.703	0.703	0.428	0.462	0.840	0.979	-	0.977	0.665
IFNGR2	0.462	0.882	0.115	0.840	0.935	0.132	0.926	-	0.795	0.582
IKBKB	-	0.243	0.360	-	-	0.096	0.141	-	0.285	-
IKBKE	-	0.243	0.935	-	-	0.002	0.489	-	0.743	-
IKBKG	-	0.665	0.935	-	-	0.935	0.567	-	0.651	-
IRF1	0.360	0.389	0.132	0.132	0.197	0.703	0.059	-	0.651	-
IRF2	-	0.096	0.001	-	0.977	0.882	0.260	-	0.701	0.267
IRF3	-	0.267	0.028	-	-	0.977	0.071	-	0.615	-
IRF5	-	0.197	0.267	-	-	0.216	0.415	-	0.795	-
IRF7	-	0.216	0.151	-	-	0.840	0.701	-	0.490	-
IRF8	-	-	-	-	0.243	-	-	-	-	-
IRF9	-	0.327	0.752	-	-	0.037	0.617	-	0.615	-
JAK1	0.132	0.582	0.267	0.171	0.151	0.935	0.036	-	0.387	0.028
JAK2	0.665	0.243	0.216	0.389	0.703	0.582	0.141	-	0.977	0.299
MYD88	-	0.037	0.428	-	-	0.011	0.701	-	0.567	-
NFKB1	-	0.882	0.327	-	-	0.327	0.047	-	0.071	-
PRKACA	0.840	0.216	0.022	0.360	0.793	0.619	0.210	-	0.701	-
PRKCD	0.665	0.840	0.360	0.327	0.132	0.151	0.701	-	0.889	0.216
PRKRA	-	0.793	0.151	-	-	0.360	0.887	-	0.490	-
RELA	-	0.665	0.058	-	-	0.752	0.071	-	0.257	-
RIPK1	-	0.046	0.058	-	-	0.197	0.088	-	0.455	-
SOCS1	0.428	0.752	0.665	0.327	0.752	0.703	0.887	0.505	0.089	0.793
SOCS3	0.428	0.935	0.197	0.462	0.015	0.299	0.036	0.209	0.387	0.002
SPI1	-	-	-	-	0.536	-	-	-	-	-
STAT1	0.793	0.151	0.015	0.115	0.501	0.752	0.059	-	0.490	0.083
STAT2	-	0.793	0.665	-	-	0.058	0.617	-	0.137	-
TBK1	-	0.151	0.058	-	-	0.299	0.047	-	0.320	-
TICAM1	-	0.752	0.132	-	-	0.793	0.260	-	0.101	-
TICAM2	-	0.840	0.935	-	-	0.501	0.102	-	0.101	-
TLR3	-	0.069	0.501	-	-	0.501	0.320	-	0.227	-
TLR7	-	0.501	0.665	-	-	0.005	0.532	-	0.929	-
TLR9	-	0.197	0.243	-	-	0.115	0.141	-	0.207	-
TP53	-	0.096	0.058	-	-	0.977	0.102	-	0.795	-
TRADD	-	0.243	0.171	-	-	0.703	0.489	-	0.795	-
TRAF2	-	0.882	0.115	-	-	0.132	0.926	-	0.795	-
TRAF6	-	0.299	0.069	-	-	0.501	0.036	-	0.490	-
TYK2	-	0.582	0.360	-	-	0.428	0.489	-	0.889	-

Table SA7: Spearman cross-correlation p-values for multiple sclerosis (MS) patients for controllable driver-target node pairs. Values for pairs that were not controllable are not reported (-). Targets IL15, PSMB9, and SP100 could not be controlled by any driver and are omitted from the table.

M0	HC 1	HC 2	HC 3	HC 4	HC 5	HC 6	HC 7	HC 8
AIFM1	8.905	8.673	8.806	8.798	8.986	8.733	8.940	8.875
APAF1	8.127	8.284	8.042	7.999	8.149	8.241	8.035	8.017
ATF2	8.102	8.206	7.941	8.301	8.158	8.219	8.135	8.115
BAK1	8.312	7.959	8.313	7.894	8.046	7.817	8.002	8.298
BAX	8.664	8.785	8.716	8.568	8.537	8.836	8.699	8.670
BBC3	6.621	6.490	6.713	6.305	6.371	6.809	6.360	6.336
BCL2	6.067	5.871	5.487	5.451	5.663	5.549	5.577	5.891
BCL2L1	8.831	8.650	8.594	8.545	8.621	8.411	8.591	8.809
BID	8.553	8.843	8.349	8.772	8.649	8.716	8.771	8.670
BIRC2	8.601	8.414	8.567	8.408	8.446	8.346	8.370	8.492
BIRC3	7.605	6.566	7.761	7.774	7.384	6.970	7.109	7.415
CASP1	8.963	9.305	9.039	9.250	8.894	9.297	9.512	9.071
CASP10	6.885	6.626	7.219	7.080	7.045	6.937	7.107	6.923
CASP2	7.721	8.026	7.882	7.950	7.764	7.747	7.692	7.745
CASP3	7.469	7.325	7.651	7.355	7.418	7.519	7.347	7.400
CASP4	8.716	9.072	8.963	9.221	8.823	9.078	9.115	8.800
CASP6	6.380	6.557	6.563	6.575	6.423	6.623	6.694	6.453
CASP7	7.431	7.448	7.595	7.442	7.444	7.224	7.430	7.491
CASP8	6.908	7.052	7.136	7.023	6.992	6.955	7.032	6.954
CASP9	7.080	6.838	7.015	6.985	7.194	6.777	6.827	7.141
CCL5	6.554	5.376	9.019	9.032	6.152	6.005	6.213	7.192
CFLAR	8.333	8.279	8.331	8.373	8.333	8.219	8.188	8.214
CHUK	8.734	8.704	8.624	8.693	8.780	8.697	8.902	8.813
CIITA	7.211	8.597	7.213	8.157	7.585	8.604	7.922	7.432
CXCL10	4.324	8.678	10.792	10.791	4.001	6.920	6.396	7.669
CXCL11	1.994	5.361	9.017	9.042	1.614	3.848	4.821	6.164
CXCL9	0.000	4.618	10.422	10.434	2.571	3.114	5.543	6.836
CYCS	10.773	10.679	10.662	10.485	10.762	10.595	10.825	10.828
DFFA	6.989	7.077	6.902	7.094	7.172	7.249	7.161	7.093
DIABLO	9.117	8.956	9.019	9.055	8.962	9.023	9.040	9.088
EIF2AK2	7.651	8.310	8.359	8.474	7.771	8.051	7.874	7.897
ENDOG	5.237	5.146	4.920	5.369	4.974	5.345	5.078	5.461
EP300	9.343	9.239	9.200	9.337	9.293	9.353	9.250	9.397
FADD	8.344	8.090	8.293	8.238	8.345	8.132	8.195	8.318
FAS	6.262	6.488	6.812	6.933	6.220	6.582	6.669	6.552
FASLG	1.217	0.000	0.000	0.000	1.304	0.000	1.378	0.000
GBP1	8.393	8.633	10.013	10.014	8.527	8.382	8.450	8.682
HLA-B	11.212	10.298	9.029	10.847	4.358	11.638	11.145	7.325
ICAM1	9.865	10.266	9.788	10.257	9.868	10.493	10.053	9.990
IFNA1	0.000	0.000	1.352	0.000	1.103	0.000	0.555	0.000
IFNAR1	9.228	8.919	9.186	9.110	9.197	8.928	9.245	9.094
IFNAR2	7.976	7.823	8.151	8.000	8.052	7.992	8.118	7.915
IFNB1	0.000	0.962	0.000	1.073	0.000	0.000	0.000	0.000
IFNG	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
IFNGR1	9.712	10.156	9.524	9.925	10.004	10.483	10.286	9.743
IFNGR2	9.468	9.551	9.518	9.807	9.561	9.667	9.749	9.521
IKBKB	8.256	8.423	8.363	8.351	8.366	8.356	8.297	8.220
IKBKE	8.601	8.705	8.569	8.738	8.685	8.655	8.548	8.767
IKBKG	7.948	8.004	7.830	8.026	7.924	7.853	7.903	7.920
IL12B	1.428	2.114	6.206	6.281	0.000	0.000	2.664	3.869
IL15	5.543	5.210	5.937	5.985	5.212	5.300	5.626	5.489
IL1B	6.621	7.531	8.790	8.797	6.276	6.501	6.794	7.316
IRF1	7.233	6.954	7.983	7.994	7.078	6.902	7.068	7.026
IRF2	7.456	7.574	7.494	7.638	7.492	7.532	7.381	7.273
IRF3	7.244	7.514	7.473	7.230	7.372	7.427	7.360	7.223
IRF4	4.292	5.445	5.993	5.990	4.546	5.538	4.062	4.744
IRF5	9.809	9.975	9.819	9.736	9.724	9.786	9.859	9.828
IRF7	5.794	7.123	7.423	7.519	5.922	6.612	5.577	6.247
IRF8	8.265	8.437	8.307	8.542	8.342	8.514	8.423	8.403
IRF9	7.767	7.972	7.882	7.944	7.632	7.818	7.645	7.742
JAK1	10.102	9.993	9.917	10.002	9.985	10.059	10.047	9.920
JAK2	8.466	8.377	8.722	8.734	8.474	8.285	8.365	8.532
MDM2	9.512	9.220	9.520	9.546	9.117	9.335	9.549	9.510
MYD88	10.918	11.245	10.748	11.002	10.962	11.123	11.061	10.322
NFKB1	9.150	9.448	9.171	9.506	9.014	9.394	9.235	9.186
NFKB2	9.122	8.870	9.078	9.114	8.808	8.652	8.676	9.074
NFKBIA	8.894	9.422	9.105	9.490	8.948	9.261	9.161	9.107
NOS2	0.000	0.000	0.889	0.000	0.000	0.000	0.000	0.000
OAS1	7.541	8.102	8.822	8.776	7.849	8.002	7.805	7.524
PARP1	8.682	8.816	8.479	8.882	8.425	8.607	8.559	8.508
PLSCR1	8.923	9.430	9.304	9.365	9.054	9.388	9.102	9.146
PRKACA	8.796	9.172	8.751	9.017	8.817	9.084	9.019	8.828
PRKCD	10.392	10.376	10.214	10.307	10.383	10.164	10.354	10.375
PRKRA	6.953	6.895	6.947	6.842	7.011	6.869	6.983	6.969
PSMB9	8.163	8.054	8.557	8.382	7.990	7.770	8.089	8.119
RELA	8.171	8.155	8.200	8.416	8.031	8.044	8.044	8.103
RELB	7.822	8.005	7.699	8.034	7.624	7.969	7.660	7.825
RIPK1	7.931	7.844	7.988	8.074	7.764	7.890	7.805	7.788
SOCS1	0.584	0.593	1.794	1.073	0.000	0.541	0.555	0.000
SOCS3	9.126	8.404	8.736	8.423	9.041	8.509	8.539	8.719
SOD2	10.051	9.952	10.284	10.725	9.817	9.910	10.236	9.475
SP100	8.473	8.590	8.622	8.760	8.382	8.604	8.464	8.422
SPI1	10.819	10.926	10.741	10.873	10.847	10.936	10.821	10.863
STAT1	9.942	9.989	10.660	10.692	10.041	9.866	10.046	10.058
STAT2	7.201	7.616	7.686	7.779	7.285	7.479	7.399	7.275
TBK1	8.621	8.559	8.668	8.672	8.660	8.490	8.504	8.542
TICAM1	8.515	8.606	8.544	8.614	8.246	8.546	8.232	8.536
TICAM2	8.641	8.833	8.668	8.827	8.655	8.882	8.735	8.692
TLR3	4.324	5.638	6.217	6.423	5.028	5.424	5.330	4.247
TLR7	6.668	8.073	7.613	7.631	6.495	8.035	7.235	6.941
TLR9	9.757	9.562	9.636	9.809	9.882	9.727	9.723	9.756
TNFRSF10A	7.335	7.034	7.109	6.550	7.312	6.661	6.779	6.684
TNFRSF10B	7.626	7.808	7.614	7.730	7.424	7.709	7.654	7.859
TNFSF10	5.363	8.369	9.148	9.136	6.083	7.515	6.084	6.469
TP53	6.919	7.319	7.034	7.381	7.033	7.326	7.224	7.132
TRADD	5.908	5.954	6.001	5.785	5.828	6.088	5.766	6.001
TRAF1	7.253	6.321	6.981	6.823	6.677	6.055	5.763	6.689
TRAF2	7.233	7.010	7.088	7.216	7.164	6.987	7.073	7.306
TRAF6	6.487	6.259	6.343	6.236	6.315	6.372	6.222	6.343
TYK2	9.089	9.265	9.118	9.083	9.004	9.144	8.862	9.106
XIAP	8.123	7.712	7.867	7.850	8.119	7.848	8.059	8.137

Table SA8: Log-transformed gene expression for healthy controls (HC), in the 'alert' (M0) macrophage activation state.

M1	HC 1	HC 2	HC 3	HC 4	HC 5	HC 6	HC 7	HC 8
AIFM1	8.615	7.912	8.496	8.189	8.604	7.933	8.398	8.430
APAF1	8.134	8.186	8.132	8.154	8.382	8.627	8.507	8.296
ATF2	7.945	8.737	7.993	8.634	8.339	8.801	8.550	8.343
BAK1	9.053	9.183	8.969	8.673	8.897	8.774	8.964	9.045
BAX	8.614	9.429	8.808	8.968	8.681	9.187	9.035	9.057
BBC3	7.053	8.465	7.422	7.664	7.465	8.407	7.976	7.747
BCL2	5.342	5.820	4.942	5.617	5.191	5.225	5.716	5.372
BCL2L1	8.735	8.938	8.645	8.724	8.591	8.630	8.537	8.784
BID	8.673	9.212	8.880	9.267	9.110	9.040	9.356	9.042
BIRC2	8.675	9.358	8.774	8.988	8.804	8.963	9.113	8.941
BIRC3	9.389	11.345	9.550	10.864	10.285	9.823	10.476	10.344
CASP1	9.839	10.421	9.934	10.222	10.157	10.694	10.556	10.148
CASP10	8.007	8.085	8.269	8.251	8.540	8.199	8.671	8.394
CASP2	7.627	7.802	7.537	7.741	7.795	7.634	7.707	7.734
CASP3	8.121	9.214	8.412	8.368	8.731	8.587	8.537	8.358
CASP4	9.783	10.589	10.008	10.311	10.156	10.936	10.631	10.153
CASP6	6.340	6.624	6.463	6.489	6.656	6.265	6.620	6.509
CASP7	8.195	9.605	8.446	8.952	8.864	9.144	9.152	8.705
CASP8	7.462	7.797	7.426	7.473	7.651	7.668	7.659	7.542
CASP9	7.379	7.247	7.445	7.097	7.055	6.957	7.091	7.299
CCL5	11.156	12.785	11.591	12.279	11.382	12.897	12.534	11.795
CFLAR	9.491	10.691	9.655	10.195	10.124	10.271	10.425	9.977
CHUK	9.013	8.961	8.882	8.835	9.103	9.143	9.148	9.046
CIITA	8.812	9.291	8.626	9.379	9.096	9.390	9.307	9.041
CXCL10	11.964	14.233	12.714	13.661	13.592	14.033	14.225	13.487
CXCL11	9.803	13.120	11.015	12.314	12.033	12.457	12.861	12.067
CXCL9	12.872	13.742	13.088	13.765	13.150	14.170	13.569	13.541
CYCS	10.436	10.027	10.379	9.963	10.432	9.759	10.156	10.411
DFFA	6.851	7.016	6.807	7.005	7.064	6.987	7.045	6.954
DIABLO	9.261	9.159	9.147	9.172	9.177	9.244	9.248	9.191
EIF2AK2	9.425	10.436	9.757	9.967	10.075	9.942	10.421	9.958
ENDOG	4.881	4.169	4.730	4.778	4.489	4.231	4.191	4.572
EP300	9.535	9.582	9.496	9.572	9.668	9.657	9.559	9.595
FADD	8.506	8.916	8.392	8.626	8.689	8.723	8.794	8.596
FAS	6.854	9.182	7.539	8.416	7.671	9.190	8.487	8.079
FASLG	2.355	1.523	1.494	2.774	2.551	0.000	4.869	0.713
GBP1	12.238	13.082	12.494	12.904	12.610	13.314	13.149	12.747
HLA-B	11.931	11.876	9.179	11.932	5.275	13.018	12.093	7.755
ICAM1	11.050	11.863	11.072	11.568	11.105	12.021	11.667	11.268
IFNA1	1.351	2.205	2.806	0.000	2.443	1.682	3.080	3.216
IFNAR1	9.393	9.386	9.533	9.270	9.621	9.609	9.521	9.504
IFNAR2	8.401	8.754	8.528	8.516	8.727	8.841	8.821	8.434
IFNB1	1.067	1.027	1.278	0.000	0.000	1.454	2.250	1.980
IFNG	0.000	2.101	0.000	4.186	0.000	3.382	5.499	0.000
IFNGR1	8.797	9.648	9.125	9.455	9.140	10.166	9.787	9.212
IFNGR2	9.677	10.215	9.826	10.126	10.102	10.310	10.280	9.984
IKBKB	8.472	8.945	8.560	8.800	8.597	8.678	8.662	8.575
IKBKE	9.002	9.390	9.071	9.224	9.316	9.168	9.178	9.284
IKBKG	8.261	8.398	8.124	8.303	8.171	8.323	8.226	8.126
IL12B	6.411	10.248	7.494	9.974	7.685	7.385	10.435	9.724
IL15	8.313	8.676	8.145	8.030	8.591	8.771	8.541	8.388
IL1B	10.393	12.023	10.875	12.141	10.135	12.586	12.451	10.999
IRF1	10.331	10.758	10.328	10.631	10.411	10.996	10.744	10.495
IRF2	7.896	8.919	8.142	8.572	8.476	8.833	8.631	8.281
IRF3	7.967	7.844	7.980	7.938	7.975	7.794	7.971	7.889
IRF4	7.256	9.826	7.569	9.531	8.217	8.257	9.155	8.257
IRF5	10.268	10.698	10.219	10.205	10.177	9.915	10.301	10.358
IRF7	9.010	10.175	9.350	9.847	9.384	9.860	9.382	9.174
IRF8	8.762	9.621	9.073	9.590	9.119	9.354	9.538	9.310
IRF9	8.659	8.952	8.657	8.883	8.699	8.729	8.858	8.773
JAK1	10.210	10.906	10.243	10.761	10.388	10.717	10.672	10.466
JAK2	10.263	10.297	10.215	10.313	10.527	10.472	10.364	10.556
MDM2	9.903	10.854	10.129	10.476	10.203	10.520	10.598	10.697
MYD88	11.025	11.018	10.986	10.994	11.178	11.167	11.239	10.574
NFKB1	9.771	11.404	10.065	11.028	10.312	11.012	11.088	10.436
NFKB2	9.901	11.207	10.122	10.851	10.221	10.861	10.778	10.396
NFKBIA	10.017	11.633	10.306	11.283	10.496	11.667	11.259	10.792
NOS2	0.000	2.464	1.003	0.000	2.443	2.857	1.060	1.980
OAS1	9.868	11.017	10.296	10.511	10.695	10.624	10.914	10.442
PARP1	8.121	8.272	8.058	8.334	8.288	8.263	8.255	8.239
PLSCR1	10.112	11.014	10.460	10.638	10.573	11.218	10.960	10.498
PRKACA	8.435	8.217	8.511	8.516	8.262	8.648	8.472	8.424
PRKCD	10.463	10.719	10.481	10.675	10.582	10.708	10.642	10.677
PRKRA	6.691	6.484	6.640	6.571	6.611	6.445	6.494	6.595
PSMB9	10.336	10.679	10.322	10.274	10.322	10.554	10.533	10.256
RELA	8.772	9.818	8.988	9.624	9.010	9.640	9.490	9.137
RELB	8.393	9.704	8.638	9.300	8.680	9.448	9.062	8.882
RIPK1	8.535	9.263	8.616	8.950	8.641	9.036	8.991	8.666
SOCS1	3.322	4.369	3.324	4.254	3.560	4.419	3.894	3.594
SOCS3	10.813	11.001	10.406	10.420	10.269	11.389	10.748	10.363
SOD2	12.142	12.956	12.446	12.876	12.539	13.222	13.149	12.474
SP100	9.256	10.098	9.440	9.741	9.625	9.960	9.905	9.636
SPI1	10.787	10.751	10.758	10.807	10.851	10.871	10.595	10.731
STAT1	12.610	12.253	12.632	12.308	12.685	12.337	12.440	12.563
STAT2	9.042	9.442	9.363	9.318	9.541	9.316	9.586	9.377
TBK1	9.271	9.783	9.357	9.508	9.573	9.975	9.810	9.512
TICAM1	9.197	9.705	9.174	9.416	9.189	9.139	9.283	9.277
TICAM2	9.120	9.727	9.211	9.534	9.410	9.937	9.654	9.398
TLR3	5.566	7.439	6.809	7.187	7.051	6.322	7.213	6.757
TLR7	7.254	7.567	8.002	7.622	8.464	7.550	7.831	8.418
TLR9	9.765	9.477	9.718	9.775	9.906	9.587	9.582	9.757
TNFRSF10A	8.508	8.543	8.136	7.733	8.327	8.273	8.267	7.793
TNFRSF10B	6.938	8.124	7.545	7.951	7.037	8.246	7.987	7.895
TNFSF10	10.245	11.932	10.954	11.377	11.513	11.587	11.766	11.234
TP53	6.696	6.907	6.909	7.074	7.273	6.797	6.961	7.162
TRADD	6.833	7.745	6.877	7.334	7.084	7.556	7.178	6.970
TRAF1	8.632	10.303	8.590	9.677	9.031	9.706	10.080	9.274
TRAF2	8.075	8.320	7.872	8.006	8.222	8.189	8.130	8.085
TRAF6	6.712	7.576	6.730	7.273	7.068	7.459	7.191	7.022
TYK2	9.658	9.260	9.571	9.383	9.330	9.264	9.155	9.387
XIAP	8.197	7.905	8.160	7.719	8.153	8.187	8.198	8.144

Table SA9: Log-transformed gene expression for healthy controls (HC), in the 'pro-infalammatory' (M1) macrophage activation state.

M0	MS 1	MS 2	MS 3	MS 4	MS 5	MS 6	MS 7	MS 8
AIFM1	8.945	8.760	8.623	8.821	8.813	8.844	8.881	9.014
APAF1	7.751	8.232	8.147	7.699	7.757	8.110	7.877	7.585
ATF2	8.325	8.163	8.760	8.455	8.493	8.456	8.795	8.833
BAK1	8.076	7.922	7.808	7.994	8.199	8.031	7.954	8.397
BAX	8.264	8.578	8.456	8.741	8.836	8.661	8.923	8.489
BBC3	6.975	6.251	6.297	6.859	6.444	6.372	6.544	6.682
BCL2	5.709	5.572	5.529	5.663	5.711	5.827	5.215	5.671
BCL2L1	8.520	8.278	8.253	8.395	8.433	8.093	8.261	8.684
BID	8.375	8.037	7.533	8.391	7.813	8.014	7.886	8.547
BIRC2	8.273	8.457	8.396	8.038	8.232	8.136	8.334	8.500
BIRC3	7.740	6.507	5.240	6.495	5.266	6.242	5.788	7.364
CASP1	8.584	9.309	9.256	9.195	9.028	9.001	9.274	8.878
CASP10	6.542	7.054	6.682	6.664	6.880	6.702	7.064	6.403
CASP2	7.656	8.072	8.342	7.936	7.885	7.991	7.716	7.631
CASP3	7.192	7.635	7.039	6.799	6.830	6.838	7.026	6.969
CASP4	9.047	8.940	9.313	9.446	9.269	9.337	9.410	9.252
CASP6	6.573	6.744	7.086	6.679	7.175	7.068	6.922	7.205
CASP7	7.309	7.302	7.192	7.314	7.279	6.935	7.098	7.409
CASP8	6.869	7.229	7.013	6.782	7.057	7.008	7.221	6.858
CASP9	7.080	7.150	6.372	6.959	6.726	6.530	6.913	7.443
CCL5	7.329	5.547	5.463	7.867	6.215	7.368	4.360	8.911
CFLAR	8.343	8.063	7.969	8.075	7.836	7.886	8.075	8.138
CHUK	8.333	8.562	8.113	8.261	8.141	8.041	8.229	8.472
CIITA	7.922	7.841	8.436	8.860	8.086	8.954	9.311	7.804
CXCL10	3.325	2.945	4.194	6.927	3.527	6.803	5.782	4.347
CXCL11	2.402	0.970	2.401	4.546	0.945	3.486	2.413	1.983
CXCL9	2.777	2.945	3.974	5.663	3.378	4.887	3.450	5.112
CYCS	10.139	10.180	9.493	9.987	9.742	9.719	9.833	10.194
DFFA	7.379	7.008	7.438	7.352	7.598	7.391	7.600	7.482
DIABLO	9.079	8.957	9.002	8.926	8.874	8.787	8.945	9.160
EIF2AK2	7.764	7.873	8.174	8.466	8.049	8.581	8.212	7.909
ENDOG	5.490	4.630	4.766	5.794	5.700	5.656	4.951	5.054
EP300	9.384	9.157	9.457	9.333	9.324	9.426	9.652	9.493
FADD	8.154	8.062	7.995	8.034	8.046	7.793	7.664	8.212
FAS	5.889	6.486	6.097	6.285	6.232	5.914	6.597	6.389
FASLG	0.984	0.970	0.000	2.842	0.000	0.981	2.212	0.000
GBP1	8.610	8.156	8.120	8.571	8.370	8.635	8.773	8.817
HLA-B	11.343	9.214	11.067	11.288	11.605	9.089	11.700	4.217
ICAM1	9.845	9.469	9.228	10.285	9.410	10.053	9.289	9.807
IFNA1	0.000	2.021	0.000	0.000	0.945	0.000	2.580	0.000
IFNAR1	8.905	9.138	8.667	8.505	8.646	8.283	8.277	8.635
IFNAR2	8.118	8.073	8.071	7.924	8.337	8.000	8.072	8.096
IFNB1	0.000	0.000	0.000	0.000	0.000	1.252	0.000	0.000
IFNG	0.000	0.000	0.000	1.027	0.000	0.000	0.000	0.000
IFNGR1	9.401	10.078	9.725	9.830	9.657	10.041	9.935	9.584
IFNGR2	9.738	9.839	10.012	9.810	9.771	9.713	9.507	9.870
IKBKB	8.484	8.429	8.563	8.536	8.396	8.548	8.551	8.213
IKBKE	8.668	8.203	7.919	8.767	8.192	8.411	8.237	8.979
IKBKG	8.249	7.558	7.953	8.283	7.991	8.022	7.934	7.887
IL12B	0.000	0.970	0.000	0.000	1.422	1.466	1.960	3.374
IL15	4.942	4.864	5.122	5.128	5.336	5.216	5.263	5.432
IL1B	7.562	6.378	6.155	7.766	5.962	7.178	6.087	6.870
IRF1	7.744	7.161	6.963	7.388	7.355	7.078	7.433	7.541
IRF2	7.696	7.526	7.780	7.623	7.563	7.510	7.697	7.669
IRF3	7.510	7.576	7.420	7.674	7.551	7.708	7.551	7.343
IRF4	4.936	4.291	5.122	5.028	3.735	6.220	5.315	5.131
IRF5	10.234	9.715	9.897	9.960	9.871	9.811	10.118	10.021
IRF7	6.506	5.491	6.258	7.633	6.495	7.792	6.872	6.445
IRF8	8.403	8.045	8.089	8.342	8.351	8.185	8.323	8.240
IRF9	7.860	7.625	7.725	8.023	7.649	7.875	7.915	7.610
JAK1	10.015	9.965	9.932	9.822	9.784	9.732	9.951	9.759
JAK2	8.461	8.484	8.034	8.169	8.021	8.393	8.300	8.346
MDM2	9.056	9.326	8.886	9.221	9.065	8.865	9.767	9.515
MYD88	11.007	10.799	10.982	11.229	10.798	11.164	11.379	11.402
NFKB1	8.836	8.829	8.585	9.169	8.780	9.122	8.916	9.008
NFKB2	9.335	8.660	8.045	9.032	8.373	8.850	7.992	9.103
NFKBIA	9.368	8.760	8.881	9.626	8.913	9.353	9.247	9.772
NOS2	0.000	1.453	1.101	1.027	0.945	0.000	0.000	0.000
OAS1	8.055	7.953	7.740	8.554	7.803	8.187	7.896	7.764
PARP1	8.157	8.549	8.962	8.753	9.128	8.761	8.593	8.043
PLSCR1	8.678	8.996	9.030	9.261	8.901	9.291	9.100	9.189
PRKACA	8.854	8.902	9.006	9.202	8.974	9.104	9.136	8.878
PRKCD	10.671	10.297	10.200	10.430	10.208	10.267	10.342	10.534
PRKRA	6.570	6.930	6.854	6.596	6.716	6.611	6.669	6.568
PSMB9	8.662	7.873	8.275	8.577	8.562	8.899	8.362	8.438
RELA	8.712	8.052	8.951	8.836	8.780	8.838	9.007	8.943
RELB	7.768	6.870	6.556	7.707	6.680	7.221	6.607	7.769
RIPK1	8.063	7.989	8.163	8.040	8.068	7.881	8.140	8.044
SOCS1	0.608	0.000	0.000	0.640	0.000	1.921	0.701	0.715
SOCS3	9.059	8.540	7.783	8.766	8.052	8.306	7.712	9.148
SOD2	10.224	9.102	8.650	9.886	8.712	9.133	9.040	10.282
SP100	8.427	8.636	8.613	8.549	8.430	8.587	8.606	8.452
SPI1	11.371	10.573	11.177	11.473	11.286	11.361	11.477	11.426
STAT1	9.910	9.759	9.290	10.016	9.499	9.729	9.673	9.778
STAT2	7.512	7.210	7.210	7.753	7.438	7.939	7.653	6.887
TBK1	8.596	8.508	8.661	8.323	8.526	8.318	8.632	8.508
TICAM1	8.851	7.904	8.555	8.776	8.327	8.733	8.639	8.891
TICAM2	8.419	8.702	8.671	8.520	8.589	8.588	8.665	8.355
TLR3	4.595	4.845	4.081	4.323	4.817	5.136	4.770	4.725
TLR7	5.956	7.540	8.096	7.607	8.429	8.024	8.625	7.297
TLR9	9.722	9.387	9.163	9.784	9.351	9.348	9.233	9.682
TNFRSF10A	6.972	6.365	5.978	6.737	6.874	6.844	6.062	7.242
TNFRSF10B	7.337	7.375	7.267	7.875	7.714	7.527	7.849	7.361
TNFSF10	6.467	7.168	8.032	8.030	7.627	8.749	8.183	7.017
TP53	7.162	6.983	7.354	7.519	7.199	7.335	7.130	7.110
TRADD	6.105	5.322	5.821	6.266	5.964	6.077	5.538	5.576
TRAF1	6.586	5.797	4.361	6.126	4.914	5.987	4.360	6.170
TRAF2	7.484	6.803	6.578	7.069	6.804	6.854	6.330	7.113
TRAF6	6.409	6.230	6.582	6.249	6.292	6.284	6.542	6.578
TYK2	9.360	9.088	9.355	9.545	9.577	9.561	9.388	8.886
XIAP	8.070	7.988	7.761	7.674	7.847	7.851	7.858	8.231

Table SA10: Log-transformed gene expression for multiple sclerosis (MS) patients, in the 'alert' (M0) macrophage activation state.

M1	MS 1	MS 2	MS 3	MS 4	MS 5	MS 6	MS 7	MS 8
AIFM1	8.601	8.526	8.354	8.205	8.233	8.121	8.125	8.595
APAF1	8.021	8.310	8.487	8.082	7.805	8.110	7.803	8.221
ATF2	8.630	8.101	8.563	8.882	8.839	9.046	9.183	9.738
BAK1	8.892	8.817	8.830	9.262	9.104	9.300	9.069	9.051
BAX	8.396	8.780	8.968	9.103	9.105	9.021	9.199	8.924
BBC3	7.612	7.603	7.742	8.519	8.134	8.020	8.159	8.772
BCL2	5.089	5.237	5.846	6.099	5.553	5.860	5.879	6.532
BCL2L1	8.538	8.532	8.698	8.729	8.634	8.575	8.385	10.140
BID	8.695	8.986	9.362	8.963	9.029	8.627	8.781	9.042
BIRC2	8.679	8.807	8.900	8.937	8.686	9.038	8.690	10.100
BIRC3	10.206	10.020	10.034	10.364	10.175	10.546	10.181	11.783
CASP1	10.015	10.171	10.258	10.429	9.754	10.317	10.220	10.422
CASP10	7.968	8.328	8.216	8.222	7.625	7.712	7.717	7.792
CASP2	7.702	7.709	7.694	7.519	7.406	7.428	7.829	7.125
CASP3	8.298	8.411	8.066	8.346	7.486	8.176	7.673	7.790
CASP4	10.421	10.016	10.362	10.868	10.476	10.936	10.984	10.868
CASP6	6.511	6.434	6.020	6.630	6.448	6.672	6.466	6.633
CASP7	8.499	8.224	8.502	9.273	8.297	9.034	8.918	9.666
CASP8	7.334	7.541	7.501	7.405	7.317	7.348	7.349	7.395
CASP9	7.122	7.289	7.105	7.198	7.250	7.037	7.058	7.238
CCL5	11.961	11.362	12.029	13.159	12.973	13.426	13.135	14.435
CFLAR	10.012	9.818	10.027	10.401	9.778	10.235	10.173	11.121
CHUK	8.683	9.003	9.074	8.476	8.382	8.395	8.614	8.548
CIITA	9.609	9.063	9.321	9.458	9.427	9.166	10.092	9.019
CXCL10	12.600	13.330	13.587	13.983	11.805	13.315	13.661	12.497
CXCL11	10.645	11.651	11.974	12.602	10.099	11.720	12.249	11.121
CXCL9	12.982	13.714	14.126	13.580	13.077	13.890	14.763	13.770
CYCS	9.747	10.347	10.326	9.482	9.366	9.256	9.333	9.843
DFFA	7.319	6.910	6.814	7.241	7.130	7.304	7.403	7.112
DIABLO	9.172	9.162	9.233	9.080	9.022	9.091	9.191	9.105
EIF2AK2	10.112	9.712	9.640	10.553	9.707	10.302	10.114	10.654
ENDOG	5.192	4.747	4.115	5.151	5.065	4.787	3.913	4.248
EP300	9.708	9.311	9.646	9.632	9.697	9.709	9.791	10.183
FADD	8.385	8.389	8.673	8.650	8.494	8.408	8.490	8.323
FAS	7.142	8.048	8.614	8.398	7.956	8.422	8.584	9.120
FASLG	1.083	2.760	2.498	4.940	2.213	4.158	2.451	2.551
GBP1	12.977	12.679	12.785	13.470	13.040	13.294	13.686	14.123
HLA-B	12.015	10.250	12.332	12.443	12.653	10.540	12.770	4.676
ICAM1	11.233	11.399	11.699	11.797	11.955	11.919	11.868	12.134
IFNA1	0.000	1.136	0.000	0.000	0.000	1.887	2.828	6.593
IFNAR1	9.238	9.671	9.624	8.947	9.039	8.972	8.888	8.683
IFNAR2	8.902	8.566	8.623	8.649	8.764	8.678	8.790	8.585
IFNB1	0.000	0.000	1.173	1.186	0.000	0.000	2.451	7.035
IFNG	1.591	2.247	3.821	4.616	0.000	5.043	0.000	4.380
IFNGR1	8.889	9.207	9.588	9.180	9.365	9.482	9.116	9.138
IFNGR2	10.046	10.211	10.396	10.304	10.164	10.212	10.446	10.026
IKBKB	8.906	8.637	8.589	8.845	8.601	8.766	8.797	8.738
IKBKE	9.351	9.060	8.991	9.542	9.401	9.393	9.485	9.126
IKBKG	8.567	8.042	8.112	8.689	8.458	8.461	8.274	8.347
IL12B	7.528	8.880	9.534	9.322	8.669	8.726	9.893	13.766
IL15	8.040	8.246	8.377	8.159	7.812	8.000	7.829	8.707
IL1B	11.425	10.869	12.054	12.064	11.794	12.199	12.254	14.361
IRF1	10.862	10.715	10.658	11.052	10.964	11.108	11.052	11.502
IRF2	8.578	8.287	8.519	8.881	8.488	8.729	8.702	8.638
IRF3	8.159	8.188	8.005	8.042	8.182	8.026	7.926	7.715
IRF4	8.724	7.847	8.202	9.472	8.238	9.261	9.257	10.106
IRF5	10.620	10.114	10.100	10.454	10.259	10.295	10.379	10.524
IRF7	9.810	9.007	9.135	10.499	9.850	10.427	9.887	10.381
IRF8	9.055	9.457	9.454	9.697	9.264	9.395	9.592	10.809
IRF9	8.752	8.549	8.522	9.002	8.515	8.764	8.719	8.901
JAK1	10.368	10.439	10.720	10.602	10.212	10.676	10.676	10.713
JAK2	10.321	10.551	10.379	10.190	9.915	10.216	10.278	10.141
MDM2	9.845	10.066	10.393	10.411	10.392	10.285	10.586	10.524
MYD88	11.305	10.945	10.804	11.508	10.846	11.363	11.386	11.551
NFKB1	10.124	10.136	10.508	11.027	10.217	11.017	10.880	11.954
NFKB2	10.480	10.368	10.572	11.151	10.737	11.274	11.098	12.156
NFKBIA	10.904	10.805	11.075	11.733	11.391	12.081	12.198	13.383
NOS2	0.000	0.000	2.667	2.312	1.861	2.240	2.657	3.121
OAS1	10.522	10.198	10.185	11.014	10.229	10.694	10.584	10.817
PARP1	8.179	8.187	8.260	8.240	7.973	8.060	8.237	7.566
PLSCR1	10.460	10.456	10.614	10.899	10.470	10.939	10.691	11.206
PRKACA	8.497	8.527	8.793	8.403	8.712	8.403	8.506	8.340
PRKCD	10.862	10.686	10.719	10.876	10.810	10.829	10.846	11.077
PRKRA	6.187	6.764	6.879	6.329	6.331	6.181	6.178	6.448
PSMB9	10.566	10.318	10.542	10.944	10.906	11.257	11.041	10.612
RELA	9.743	9.121	9.400	10.308	10.158	10.446	10.708	11.271
RELB	8.768	8.468	8.906	9.253	9.083	9.373	9.072	9.404
RIPK1	8.958	8.567	8.541	9.161	8.670	8.971	8.934	9.611
SOCS1	4.151	3.718	3.284	5.145	4.150	5.121	4.673	5.115
SOCS3	10.435	10.576	10.749	10.812	10.749	11.742	10.876	12.889
SOD2	12.625	12.612	12.714	12.907	12.607	13.168	13.260	13.854
SP100	9.530	9.596	9.714	9.926	9.444	9.760	9.679	9.816
SPI1	11.373	10.468	10.558	11.376	11.468	11.401	11.322	11.371
STAT1	12.602	12.604	12.330	12.293	12.081	11.965	12.012	11.978
STAT2	9.639	9.240	9.086	9.802	9.298	9.719	9.570	9.806
TBK1	9.589	9.464	9.566	9.587	9.442	9.558	9.649	10.053
TICAM1	9.635	8.804	9.026	9.834	9.378	9.721	9.787	11.242
TICAM2	9.260	9.352	9.568	9.321	9.061	9.329	9.475	9.356
TLR3	6.061	5.750	5.530	6.802	6.700	6.754	5.368	5.410
TLR7	7.173	8.131	7.692	8.189	6.819	7.727	7.296	8.015
TLR9	9.612	9.561	9.728	9.632	9.833	9.398	9.277	8.955
TNFRSF10A	8.090	7.737	7.591	8.432	8.120	8.589	7.864	8.572
TNFRSF10B	6.938	7.558	8.135	8.056	8.155	7.963	8.111	8.056
TNFSF10	11.269	11.146	11.274	12.170	11.233	11.893	11.412	11.157
TP53	7.167	7.098	7.104	7.141	7.084	6.698	7.130	6.612
TRADD	7.117	6.877	7.069	7.896	7.252	7.817	6.763	7.162
TRAF1	8.603	8.977	9.088	9.643	8.712	9.565	9.301	11.014
TRAF2	8.372	8.047	7.840	8.417	8.109	8.290	8.185	7.784
TRAF6	7.062	6.932	7.021	7.321	6.927	7.357	7.373	7.845
TYK2	9.644	9.532	9.349	9.676	9.958	9.558	9.487	9.269
XIAP	8.161	8.219	8.124	8.003	8.079	8.353	8.026	9.043

Table SA11: Log-transformed gene expression for multiple sclerosis (MS) patients, in the 'pro-inflammatory' (M1) macrophage activation state.

Gene	Delta	Gene (cont.)	Delta (cont.)
AIFM1	0.025	IL15	0.089
APAF1	0.019	IL1B	1.261
ATF2	0.067	IRF1	0.349
BAK1	0.128	IRF2	-0.019
BAX	0.030	IRF3	-0.076
BBC3	0.243	IRF4	0.484
BCL2	0.441	IRF5	-0.061
BCL2L1	0.348	IRF7	0.192
BID	0.454	IRF8	0.452
BIRC2	0.188	IRF9	-0.046
BIRC3	1.147	JAK1	0.139
CASP1	0.053	JAK2	0.092
CASP10	-0.126	MDM2	0.091
CASP2	-0.233	MYD88	0.019
CASP3	-0.116	NFKB1	0.451
CASP4	0.018	NFKB2	0.687
CASP6	-0.417	NFKBIA	0.698
CASP7	0.127	NOS2	-0.073
CASP8	-0.184	OAS1	0.043
CASP9	0.052	PARP1	-0.140
CCL5	1.068	PLSCR1	0.191
CFLAR	0.327	PRKACA	0.016
CHUK	0.105	PRKCD	0.172
CIITA	-0.285	PRKRA	0.090
CXCL10	2.324	PSMB9	0.048
CXCL11	2.388	RELA	0.215
CXCL9	1.670	RELB	0.710
CYCS	0.308	RIPK1	-0.074
DFFA	-0.126	SOCS1	0.647
DIABLO	-0.002	SOCS3	0.694
EIF2AK2	0.021	SOD2	0.920
ENDOG	0.084	SP100	-0.022
EP300	0.025	SPI1	-0.017
FADD	0.071	STAT1	0.210
FAS	0.437	STAT2	0.162
FASLG	0.286	TBK1	0.095
GBP1	0.820	TICAM1	0.276
HLA-B	0.119	TICAM2	0.019
ICAM1	0.698	TLR3	-0.079
IFNA1	-0.859	TLR7	-0.569
IFNAR1	0.135	TLR9	0.076
IFNAR2	-0.016	TNFRSF10A	0.226
IFNB1	0.446	TNFRSF10B	0.296
IFNG	0.688	TNFSF10	-0.270
IFNGR1	0.027	TP53	-0.021
IFNGR2	-0.017	TRADD	0.130
IKBKB	-0.062	TRAF1	0.984
IKBKE	0.326	TRAF2	0.274
IKBKG	0.069	TRAF6	0.027
IL12B	2.542	TYK2	-0.078
		XIAP	0.223

Table SA12: Values of Δ for each gene (**5.1.6**). The value of the threshold at the $75th$ percentile is 0.438.

Appendix B

Supplementary material for Section 5.2

B.1 Supplementary figures

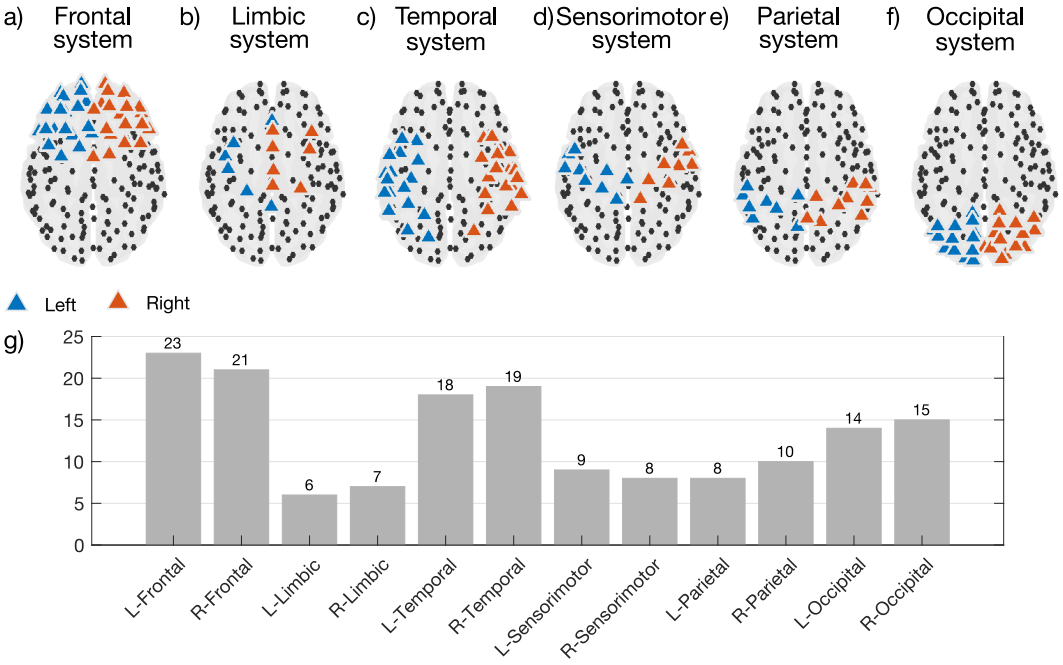


Figure SB1: **Structural systems.** Panels **a-f)** show the position of the structural systems in the connectomes, highlighting regions of interest (ROIs) in the left and right hemispheres. Panel **g)** shows the number of ROIs in each system.

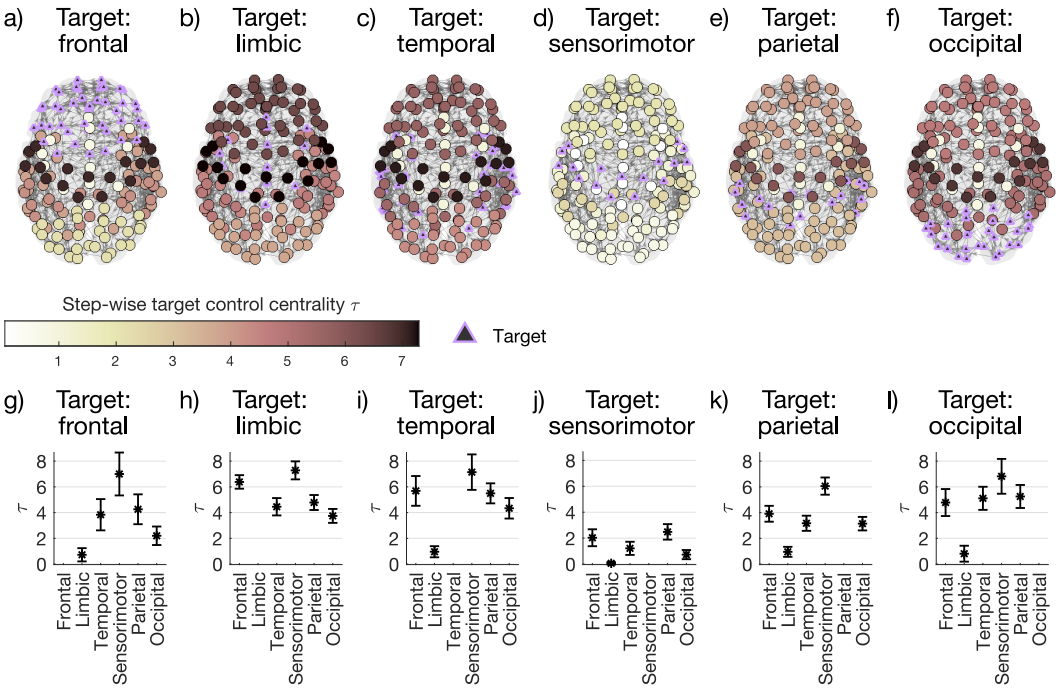


Figure SB2: **Step-wise target control centrality averaged across systems and subjects, for different target sets.** Panels **a-f)** show how the step-wise target control centrality (τ) is distributed on the connecomes at the varying of the target set. Panels **g-l)** show numerical values of the centrality for the corresponding target sets. Error bars stand for standard deviation across subjects.

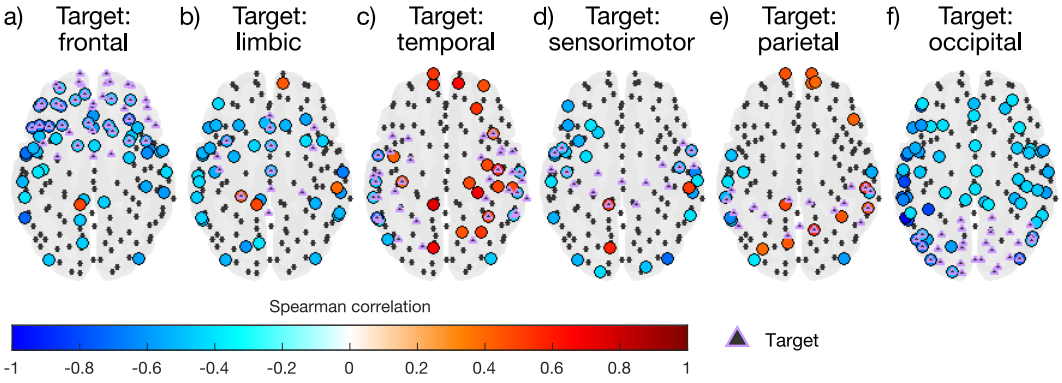


Figure SB3: **Partial correlation between step-wise target control centrality and age, corrected for outdegree**, for different target sets. Only significant ($p < 0.05$) values are shown.

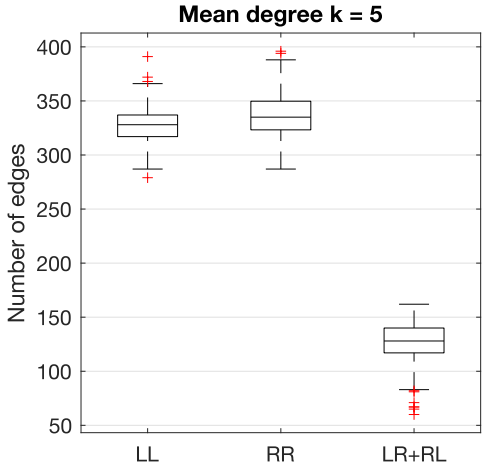


Figure SB4: **Number of edges within and between hemispheres.** As the mean degree increases, the gap between the number of edges within and between hemispheres becomes more and more accentuated. LL denotes edges within the left hemisphere, RR denotes edges within the right hemisphere, and LR+RL denotes edges between hemispheres.

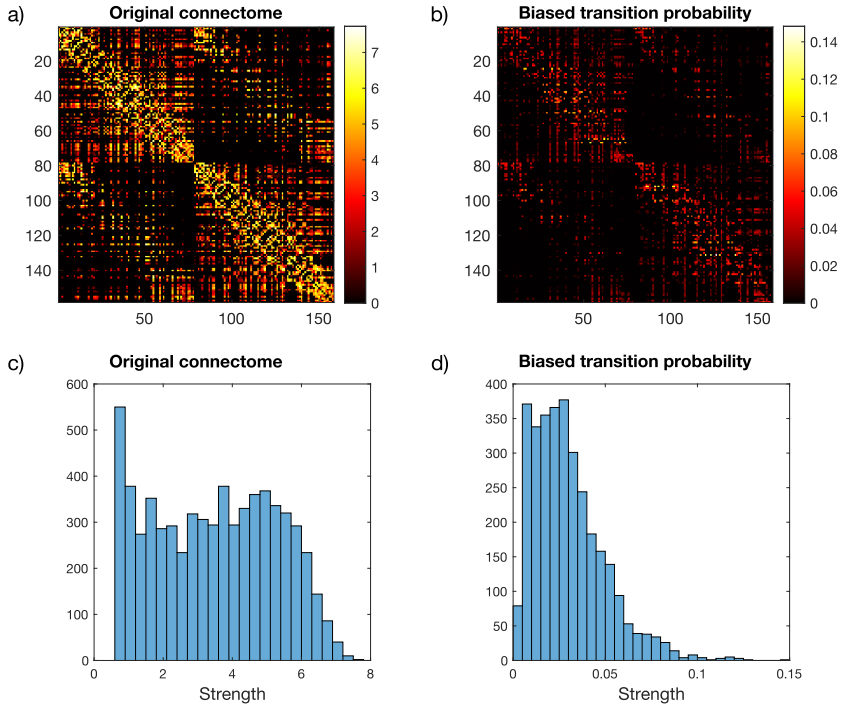
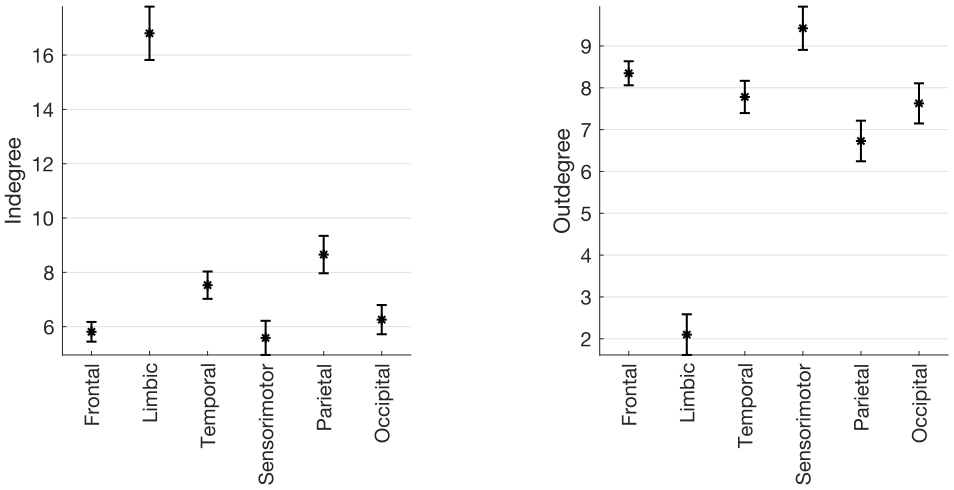


Figure SB5: **Structural connectome retrieved from the diffusion tensor imaging (DTI) data and the matrix resulting from the biased random walk process**, for an exemplary subject. Panel **a)** represents the adjacency matrix of the DTI connectome; panel **b)** the output of the biased random walk process. Panel **c)** shows a histogram for the node-strength of the original case; panel **c)** shows a histogram for the node-strength obtained from the biased random walk.



(a) Indegree of the structural systems.

(b) Outdegree of the structural systems.

Figure SB6: **Indegree and outdegree of the structural systems**, computed on the directed connectomes and averaged across mean degree k from 1 to 14, and across subjects. Error bars stand for standard deviation across subjects.

B.2 Supplementary tables

Label	Short label	System
Left Frontal Pole 1	LFP1	Frontal
Left Frontal Pole 2	LFP2	Frontal
Left Frontal Pole 3	LFP3	Frontal
Left Frontal Pole 4	LFP4	Frontal
Left Frontal Pole 5	LFP5	Frontal
Left Paracingulate 1	LPG1	Frontal
Left Frontal Pole 6	LFP6	Frontal
Left Frontal Pole 7	LFP7	Frontal
Left Frontal Pole 8	LFP8	Frontal
Left Paracingulate 2	LPG2	Frontal
Left Frontal Pole 9	LFP9	Frontal
Left Frontal Pole 10	LFP10	Frontal
Left Frontal Medial	LFMC	Frontal
Left Cingulate anterior	LCGad	Limbic
Left Superior Frontal 1	LSFG1	Frontal
Left Middle Frontal 1	LMFG1	Frontal
Left Superior Frontal 2	LSFG2	Frontal
Left Frontal Orbital 1	LFOC1	Frontal
Left Inferior Frontal pars triangularis	LIFGpt	Frontal
Left Subcallosal	LSC	Frontal
Left Frontal Orbital 2	LFOC2	Frontal
Left Temporal Pole 1	LTP1	Temporal
Left Temporal Pole 2	LTP2	Temporal
Left Insular 1	LIC1	Limbic
Left Middle Frontal 2	LMFG2	Frontal
Left Superior Frontal 3	LSFG3	Frontal
Left Precentral 1	LPGa1	Sensorimotor
Left Temporal Pole 3	LTP3	Temporal
Left Central Opercular	LCOC	Sensorimotor
Left Middle Temporal anterior	LMTGad	Temporal
Left Insular 2	LIC2	Limbic
Left Middle Frontal 3	LMFG3	Frontal
Left Temporal Fusiform anterior	LTFa	Temporal
Left Postcentral 1	LPGp1	Sensorimotor
Left Insular 3	LIC3	Limbic
Left Precentral 2	LPGa2	Sensorimotor
Left Precentral 3	LPGa3	Sensorimotor
Left Inferior Temporal posterior 1	LITGpd1	Temporal
Left Middle Temporal posterior 1	LMTGp1	Temporal
Left Temporal Fusiform posterior	LTFCpd	Temporal
Left Precentral 4	LPGa4	Sensorimotor
Left Planum Temporale 1	LPT1	Temporal
Left Postcentral 2	LPGp2	Sensorimotor
Left Supramarginal anterior	LSGad	Parietal
Left Planum Temporale 2	LPT2	Temporal
Left Postcentral 3	LPGp3	Sensorimotor
Left Parahippocampal posterior	LPGpd	Limbic
Left Middle Temporal posterior 2	LMTGpd2	Temporal
Left Inferior Temporal posterior 2	LITGpd2	Temporal
Left Parietal Operculum	LPOC	Parietal
Left Precuneous 1	LPC1	Parietal
Left Postcentral 4	LPGp4	Sensorimotor
Left Superior Parietal Lobule 1	LSPL1	Parietal
Left Temporal Occipital Fusiform 1	LTOFC1	Temporal
Left Cingulate posterior	LCGpd	Limbic
Left Supramarginal posterior	LSGpd	Parietal
Left Superior Parietal Lobule 2	LSPL2	Parietal
Left Middle Temporal temporooccipital	LMTGtp	Temporal
Left Lingual 1	LLG1	Occipital
Left Inferior Temporal temporooccipital	LITGtp	Temporal
Left Temporal Occipital Fusiform 2	LTOFC2	Temporal
Left Angular	LAG	Parietal
Left Lateral Occipital superior 1	LLOCsd1	Occipital
Left Lateral Occipital superior 2	LLOCsd2	Occipital
Left Lateral Occipital superior 3	LLOCsd3	Occipital
Left Precuneous 2	LPC2	Parietal
Left Occipital Fusiform 1	LOFG1	Temporal
Left Lateral Occipital inferior 1	LLOCid1	Occipital
Left Lingual 2	LLG2	Occipital
Left Lateral Occipital superior 4	LLOCsd4	Occipital
Left Occipital Fusiform 2	LOFG2	Temporal
Left Lateral Occipital superior 5	LLOCsd5	Occipital
Left Lateral Occipital superior 6	LLOCsd6	Occipital
Left Lateral Occipital inferior 2	LLOCid2	Occipital
Left Occipital Pole 1	LOP1	Occipital
Left Occipital Pole 2	LOP2	Occipital
Left Occipital Pole 3	LOP3	Occipital
Left Occipital Pole 4	LOP4	Occipital

Table SB1: Denomination of regions of interest (ROIs) in the left hemisphere.

Label	Short label	System
Right Frontal Pole 1	RFP1	Frontal
Right Frontal Pole 2	RFP2	Frontal
Right Frontal Pole 3	RFP3	Frontal
Right Frontal Pole 4	RFP4	Frontal
Right Frontal Pole 5	RFP5	Frontal
Right Paracingulate 1	RPG1	Frontal
Right Frontal Pole 6	RFP6	Frontal
Right Frontal Pole 7	RFP7	Frontal
Right Frontal Pole 8	RFP8	Frontal
Right Paracingulate 2	RPG2	Frontal
Right Frontal Pole 9	RFP9	Frontal
Right Frontal Pole 10	RFP10	Frontal
Right Middle Frontal 1	RMFG1	Frontal
Right Middle Frontal 2	RMFG2	Frontal
Right Inferior Frontal pars triangularis	RIFGpt	Frontal
Right Superior Frontal 1	RSFG1	Frontal
Right Cingulate anterior 1	RCGad1	Limbic
Right Insular 1	RIC1	Limbic
Right Frontal Orbital	RFOC	Frontal
Right Temporal Pole 1	RTP1	Temporal
Right Temporal Pole 2	RTP2	Temporal
Right Middle Frontal 3	RMFG3	Frontal
Right Middle Frontal 4	RMFG4	Frontal
Right Precentral 1	RPGa1	Sensorimotor
Right Cingulate anterior 2	RCGad2	Limbic
Right Insular 2	RIC2	Limbic
Right Temporal Pole 3	RTP3	Temporal
Right Central Opercular	RCOC	Sensorimotor
Right Middle Temporal anterior	RMTGad	Temporal
Right Planum Polare	RPP	Temporal
Right Superior Frontal 2	RSFG2	Frontal
Right Juxtapositional Lobule	RJL	Frontal
Right Temporal Fusiform anterior	RTFCa	Temporal
Right Precentral 2	RPGa2	Sensorimotor
Right Postcentral 1	RPGp1	Sensorimotor
Right Postcentral 2	RPGp2	Sensorimotor
Right Heschls's	RHG	Temporal
Right Cingulate posterior 1	RCGpd1	Limbic
Right Middle Temporal posterior	RMTGpd	Temporal
Right Inferior Temporal posterior 1	RITGpd1	Temporal
Right Temporal Fusiform posterior 1	RTFCpd1	Temporal
Right Superior Temporal posterior 1	RSTGpd1	Temporal
Right Precentral 3	RPGa3	Sensorimotor
Right Postcentral 3	RPG3	Sensorimotor
Right Parietal Operculum 1	RPOC1	Parietal
Right Supramarginal anterior	RSGad	Parietal
Right Cingulate posterior 2	RCGpd2	Limbic
Right Superior Temporal posterior 2	RSTGpd2	Temporal
Right Parietal Operculum 2	RPOC2	Parietal
Right Inferior Temporal posterior 2	RITGpd2	Temporal
Right Parahippocampal posterior	RPGpd	Limbic
Right Temporal Fusiform posterior 2	RTFCpd2	Temporal
Right Middle Temporal temporooccipital 1	RMTGtp1	Temporal
Right Precuneous 1	RPC1	Parietal
Right Superior Parietal Lobule 1	RSPL1	Parietal
Right Postcentral 4	RPG4	Sensorimotor
Right Supramarginal posterior	RSGp	Parietal
Right Superior Parietal Lobule 2	RSPL2	Parietal
Right Middle Temporal temporooccipital 2	RMTGtp2	Temporal
Right Inferior Temporal temporooccipital	RITGtp	Temporal
Right Temporal Occipital Fusiform	RTOFC	Temporal
Right Lingual 1	RLG1	Occipital
Right Angular	RAG	Parietal
Right Lateral Occipital superior 1	RLOCsd1	Occipital
Right Precuneous 2	RPC2	Parietal
Right Precuneous 3	RPC3	Parietal
Right Lateral Occipital superior 2	RLOCsd2	Occipital
Right Lateral Occipital superior 3	RLOCsd3	Occipital
Right Lateral Occipital inferior 1	RLOCid1	Occipital
Right Lateral Occipital inferior 2	RLOCid2	Occipital
Right Occipital Fusiform	ROFG	Temporal
Right Intracalcarine	RIC	Occipital
Right Lateral Occipital superior 4	RLOCsd4	Occipital
Right Lateral Occipital superior 5	RLOCsd5	Occipital
Right Lateral Occipital superior 6	RLOCsd6	Occipital
Right Lateral Occipital inferior 3	RLOCid3	Occipital
Right Lingual 2	RLG2	Occipital
Right Occipital Pole 1	ROP1	Occipital
Right Occipital Pole 2	ROP2	Occipital
Right Occipital Pole 3	ROP3	Occipital

Table SB2: Denomination of regions of interest (ROIs) in the right hemisphere.

System	F-values centrality	F-values self-regulation
Frontal	747	1568
Limbic	980	1453
Temporal	995	2392
Sensorimotor	694	1405
Parietal	1797	2873
Occipital	869	2179

Table SB3: F-values for the one-way anova test in the case of step-wise target control centrality and self-regulation.

rho	Frontal	Limbic	Temporal	Sensorimotor	Parietal	Occipital
LFP1	-0.015	0.148	0.206	0.030	0.175	0.038
LFP2	-0.145	-0.063	0.052	-0.099	0.008	-0.103
LFP3	-0.021	0.125	0.214	-0.019	0.111	-0.012
LFP4	-0.143	-0.047	0.081	-0.122	-0.063	-0.117
LFP5	-0.157	0.069	0.083	-0.052	0.049	-0.065
LPG1	-0.194	0.088	0.031	-0.145	0.004	-0.089
LFP6	-0.163	-0.093	0.014	-0.120	-0.009	-0.100
LFP7	-0.230	-0.166	-0.023	-0.149	-0.052	-0.156
LFP8	-0.179	-0.042	-0.003	0.052	0.064	-0.128
LPG2	-0.078	0.032	0.027	-0.125	0.007	-0.198
LFP9	-0.217	-0.143	-0.142	-0.214	-0.112	-0.279
LFP10	-0.254	-0.040	0.039	-0.032	-0.039	-0.086
LFMC	-0.106	0.046	0.040	-0.026	0.047	-0.095
LCGad	-0.236	-0.105	-0.025	-0.110	-0.057	-0.157
LSFG1	-0.203	-0.198	0.021	-0.136	-0.059	-0.133
LMFG1	-0.274	-0.203	-0.143	-0.220	-0.138	-0.238
LSFG2	-0.215	-0.213	-0.019	-0.154	-0.054	-0.156
LFOC1	-0.277	-0.141	0.046	-0.113	-0.090	-0.153
LIFGpt	-0.268	-0.222	-0.109	-0.123	-0.105	-0.232
LSC	-0.123	0.012	0.093	-0.123	0.067	-0.063
LFOC2	-0.118	-0.004	0.036	-0.178	0.004	-0.110
LTP1	-0.195	-0.140	0.030	-0.138	-0.058	-0.123
LTP2	-0.169	-0.071	-0.107	-0.134	-0.056	-0.144
LIC1	-0.026	-0.212	-0.014	-0.209	0.025	-0.123
LMFG2	-0.089	-0.041	0.038	0.013	0.069	-0.071
LSFG3	-0.173	-0.172	0.068	-0.091	0.035	-0.099
LPGa1	-0.200	-0.211	-0.106	-0.177	-0.068	-0.211
LTP3	-0.197	-0.147	-0.121	-0.180	-0.072	-0.227
LCOC	-0.276	-0.205	-0.135	-0.208	-0.145	-0.233
LMTGad	-0.257	-0.215	-0.246	-0.193	-0.194	-0.314
LIC2	-0.110	-0.109	0.164	-	0.130	-0.081
LMFG3	-0.148	-0.207	0.015	-0.163	-0.064	-0.192
LTFa	-0.066	-0.042	0.040	-0.009	0.017	-0.136
LPGp1	-0.081	-0.063	-0.055	-0.080	-0.003	-0.182
LIC3	-0.152	-0.075	-0.009	-	-0.036	-0.111
LPGa2	-0.154	-0.214	-0.064	-0.158	-0.164	-0.189
LPGa3	-0.087	-0.084	0.057	-0.097	0.005	-0.135
LITGpd1	-0.028	0.034	0.030	-0.051	0.020	-0.116
LMTGp1	-0.062	-0.167	-0.137	-0.089	-0.140	-0.245
LTFCpd	-0.098	-0.052	0.157	-0.002	0.070	-0.016
LPGa4	-0.015	-0.006	0.076	0.008	0.028	-0.143
LPT1	-0.204	-0.169	-0.180	-0.157	-0.157	-0.319
LPGp2	-0.022	0.008	0.071	0.039	0.017	-0.072
LSGad	-0.031	-0.005	0.058	-0.001	0.037	-0.103
LPT2	-0.089	0.003	-0.068	-	0.060	-0.119
LPGp3	-0.076	-0.064	0.057	-0.018	-0.068	-0.154
LPGpd	-0.032	0.166	0.102	-	0.091	-0.098
LMTGpd2	-0.087	-0.179	-0.156	-0.109	-0.198	-0.281
LITGpd2	-0.155	-0.105	-0.103	-0.118	-0.141	-0.243
LPOC	-0.131	-0.096	-0.094	-0.095	-0.112	-0.260
LPC1	-0.189	-0.218	0.021	-0.143	-0.042	-0.168
LPGp4	0.188	0.189	0.285	0.216	0.191	0.029
LSPL1	-0.105	-0.082	0.095	0.035	-0.058	-0.138
LTOFC1	-0.090	-0.090	-0.005	0.008	-0.105	-0.317
LCGpd	-0.036	-0.127	0.074	-	0.051	0.038
LSGpd	-0.083	-0.069	0.001	-0.015	0.025	-0.164
LSPL2	0.020	-0.049	0.150	0.003	-0.037	-0.069
LMTGtp	-0.284	-0.178	-0.276	-0.210	-0.249	-0.391
LLG1	-0.184	-0.102	0.118	-0.125	0.074	-0.098
LITGtp	-0.100	-0.195	-0.154	-0.072	-0.186	-0.313
LTOFC2	-0.073	0.001	0.112	-0.011	0.127	-0.130
LAG	-0.129	-0.107	-0.015	-0.063	-0.096	-0.105
LLOCsd1	0.046	0.007	0.129	0.025	-0.015	-0.100
LLOCsd2	0.036	-0.056	0.092	0.067	-0.045	-0.137
LLOCsd3	-0.060	-0.035	-0.063	-0.021	-0.035	-0.171
LPC2	-0.001	-0.031	0.085	0.022	-0.025	0.059
LOFG1	-0.080	-0.032	-0.074	-0.078	-0.124	-0.213
LLOCid1	-0.091	-0.144	-0.198	-0.052	-0.087	-0.272
LLG2	-0.179	-0.153	0.097	-0.094	0.170	-0.028
LLOCsd4	-0.061	-0.080	0.022	-0.098	0.003	-0.268
LOFG2	-0.140	-0.235	-0.052	-0.060	-0.122	-0.228
LLOCsd5	0.068	0.113	0.262	0.229	0.144	0.078
LLOCsd6	-0.053	-0.035	0.071	-0.089	0.153	-0.093
LLOCid2	-0.184	-0.202	-0.137	-0.201	-0.151	-0.249
LOP1	-0.046	-0.072	0.072	-0.011	0.034	-0.092
LOP2	-0.018	-0.083	0.052	0.005	0.010	-0.090
LOP3	-0.135	-0.133	-0.058	-0.165	-0.093	-0.195
LOP4	-0.074	-0.062	0.136	-0.075	0.026	-0.101

Table SB4: Spearman partial correlation values computed for step-wise target control centrality and age, corrected for outdegree, for regions of interest (ROIs) in the left hemisphere. Missing values correspond to cases in which the step-wise target control centrality is zero for all subjects.

pval	Frontal	Limbic	Temporal	Sensorimotor	Parietal	Occipital
LFP1	0.847	0.054	0.007	0.696	0.023	0.621
LFP2	0.059	0.415	0.497	0.199	0.915	0.180
LFP3	0.783	0.105	0.005	0.804	0.149	0.872
LFP4	0.063	0.545	0.295	0.113	0.412	0.128
LFP5	0.041	0.369	0.282	0.497	0.525	0.402
LPG1	0.011	0.252	0.686	0.059	0.962	0.249
LFP6	0.034	0.229	0.861	0.120	0.905	0.195
LFP7	0.003	0.031	0.771	0.052	0.504	0.042
LFP8	0.020	0.588	0.965	0.501	0.404	0.097
LPG2	0.313	0.676	0.722	0.105	0.929	0.009
LFP9	0.004	0.064	0.064	0.005	0.146	0.000
LFP10	0.001	0.602	0.616	0.679	0.611	0.268
LFMC	0.169	0.549	0.604	0.733	0.542	0.219
LCGad	0.002	0.175	0.750	0.152	0.459	0.041
LSFG1	0.008	0.010	0.784	0.078	0.446	0.084
LMFG1	0.000	0.008	0.063	0.004	0.072	0.002
LSFG2	0.005	0.005	0.806	0.045	0.487	0.042
LFOC1	0.000	0.067	0.547	0.141	0.245	0.046
LIFGpt	0.000	0.004	0.155	0.110	0.175	0.002
LSC	0.111	0.874	0.226	0.110	0.387	0.417
LFOC2	0.124	0.957	0.641	0.020	0.963	0.151
LTP1	0.011	0.069	0.702	0.074	0.450	0.110
LTP2	0.027	0.356	0.166	0.082	0.465	0.061
LIC1	0.740	0.005	0.858	0.006	0.751	0.111
LMFG2	0.250	0.591	0.623	0.869	0.369	0.361
LSFG3	0.024	0.025	0.378	0.238	0.648	0.200
LPGa1	0.009	0.006	0.170	0.021	0.378	0.006
LTP3	0.010	0.055	0.117	0.019	0.352	0.003
LCOC	0.000	0.007	0.080	0.006	0.059	0.002
LMTGad	0.001	0.005	0.001	0.012	0.011	0.000
LIC2	0.154	0.157	0.033	-	0.091	0.291
LMFG3	0.054	0.007	0.846	0.034	0.407	0.012
LTFa	0.390	0.582	0.601	0.909	0.825	0.078
LPGp1	0.292	0.412	0.479	0.301	0.968	0.017
LIC3	0.048	0.331	0.910	-	0.641	0.150
LPGa2	0.045	0.005	0.409	0.040	0.032	0.014
LPGa3	0.257	0.276	0.457	0.208	0.952	0.079
LITGpd1	0.713	0.659	0.695	0.512	0.797	0.133
LMTGp1	0.419	0.029	0.074	0.248	0.068	0.001
LTFCpd	0.203	0.501	0.040	0.984	0.363	0.835
LPGa4	0.850	0.938	0.324	0.915	0.719	0.062
LPT1	0.008	0.028	0.019	0.041	0.041	0.000
LPGp2	0.772	0.917	0.360	0.615	0.824	0.352
LSGad	0.688	0.952	0.450	0.992	0.636	0.183
LPT2	0.250	0.967	0.378	-	0.438	0.124
LPGp3	0.324	0.407	0.459	0.820	0.376	0.045
LPGpd	0.679	0.031	0.184	-	0.237	0.204
LMTGpd2	0.261	0.019	0.042	0.159	0.010	0.000
LITGpd2	0.043	0.172	0.180	0.124	0.066	0.001
LPOC	0.089	0.212	0.221	0.220	0.144	0.001
LPC1	0.013	0.004	0.788	0.063	0.585	0.029
LPGp4	0.014	0.014	0.000	0.005	0.012	0.703
LSPL1	0.172	0.287	0.217	0.649	0.456	0.072
LTOFC1	0.245	0.245	0.948	0.914	0.173	0.000
LCGpd	0.645	0.099	0.340	-	0.509	0.627
LSGpd	0.284	0.373	0.987	0.849	0.745	0.033
LSPL2	0.796	0.525	0.050	0.967	0.632	0.370
LMTGtp	0.000	0.020	0.000	0.006	0.001	0.000
LLG1	0.016	0.187	0.125	0.105	0.336	0.205
LITGtp	0.197	0.011	0.045	0.353	0.015	0.000
LTOFC2	0.345	0.985	0.147	0.890	0.100	0.090
LAG	0.095	0.167	0.844	0.414	0.215	0.174
LLOCsd1	0.554	0.928	0.094	0.745	0.846	0.194
LLOCsd2	0.640	0.471	0.233	0.384	0.560	0.076
LLOCsd3	0.437	0.648	0.417	0.788	0.647	0.026
LPC2	0.990	0.686	0.271	0.774	0.743	0.447
LOFG1	0.299	0.677	0.340	0.312	0.108	0.005
LLOCid1	0.240	0.061	0.010	0.498	0.257	0.000
LLG2	0.020	0.046	0.208	0.223	0.027	0.713
LLOCsd4	0.426	0.299	0.774	0.204	0.972	0.000
LOFG2	0.068	0.002	0.501	0.436	0.113	0.003
LLOCsd5	0.381	0.141	0.001	0.003	0.062	0.310
LLOCsd6	0.494	0.650	0.359	0.247	0.047	0.230
LLOCid2	0.016	0.008	0.074	0.009	0.050	0.001
LOP1	0.552	0.348	0.353	0.889	0.655	0.231
LOP2	0.812	0.281	0.501	0.945	0.898	0.245
LOP3	0.080	0.085	0.450	0.032	0.226	0.011
LOP4	0.339	0.420	0.076	0.331	0.735	0.191

Table SB5: Spearman partial correlation p-values computed for step-wise target control centrality and age, corrected for outdegree, for regions of interest (ROIs) in the left hemisphere. Missing values correspond to cases in which the step-wise target control centrality is zero for all subjects.

rho	Frontal	Limbic	Temporal	Sensorimotor	Parietal	Occipital
RFP1	-0.039	0.143	0.103	0.029	0.168	-0.006
RFP2	-0.064	0.083	0.139	-0.010	0.154	0.076
RFP3	0.100	0.170	0.251	0.075	0.159	0.079
RFP4	-0.103	0.055	0.118	0.011	0.083	-0.021
RFP5	0.125	0.146	0.179	0.007	0.087	-0.025
RPG1	-0.147	0.012	0.143	0.053	0.059	-0.026
RFP6	-0.172	-0.114	0.065	-0.078	0.027	-0.096
RFP7	-0.181	-0.003	-0.049	-0.037	0.009	-0.158
RFP8	-0.045	-0.058	-0.013	-0.091	-0.042	-0.141
RPG2	-0.138	-0.022	0.022	-0.032	0.012	-0.087
RFP9	-0.088	-0.056	0.042	-0.095	-0.001	-0.027
RFP10	-0.096	-0.047	0.210	0.038	-0.017	-0.188
RMFG1	-0.279	-0.222	0.081	-0.053	0.047	-0.061
RMFG2	-0.021	-0.012	0.118	0.101	0.159	0.061
RIFGpt	-0.126	-0.102	-0.002	-0.047	-0.018	-0.116
RSFG1	-0.209	-0.202	0.021	-0.132	0.047	-0.123
RCGad1	-0.225	-0.194	-0.045	-0.079	-0.109	-0.131
RIC1	-0.197	-0.059	0.103	-0.086	0.141	-0.155
RFOC	-0.158	-0.013	0.056	-0.049	0.074	-0.006
RTP1	-0.130	-0.062	0.156	-0.047	-0.098	-0.062
RTP2	-0.183	-0.014	-0.009	-0.039	0.064	-0.050
RMFG3	-0.165	-0.163	0.077	-0.148	0.026	-0.072
RMFG4	-0.209	-0.176	0.000	-0.105	0.024	-0.072
RPGa1	-0.072	-0.020	0.074	-0.116	0.081	-0.062
RCGad2	-0.169	-0.178	-0.034	-0.095	-0.094	-0.154
RIC2	-0.077	-0.118	0.021	-0.046	0.001	-0.183
RTP3	-0.250	-0.139	-0.108	-0.122	-0.052	-0.151
RCOC	-0.212	-0.118	-0.122	-0.194	-0.061	-0.220
RMTGad	-0.199	-0.131	-0.129	-0.131	-0.078	-0.196
RPP	-0.250	-0.050	0.069	-0.056	-0.027	-0.184
RSFG2	-0.104	-0.074	0.116	-0.052	0.089	-0.099
RJL	-0.080	-0.098	0.076	-0.087	-0.047	-0.085
RTFCa	-0.184	-0.095	-0.024	-0.062	0.016	0.021
RPGa2	-0.098	-0.090	0.186	-0.043	0.057	-0.082
RPGp1	-0.125	-0.086	-0.058	-0.101	-0.138	-0.126
RPGp2	-0.067	-0.101	-0.016	-0.192	-0.035	-0.119
RHG	-0.130	0.050	0.227	-0.091	0.054	-0.018
RCGpd1	-0.110	-0.193	0.013	-0.149	-0.137	-0.154
RMTGpd	-0.165	-0.262	-0.163	-0.248	-0.213	-0.206
RITGpd1	-0.039	0.031	0.123	0.021	0.085	0.047
RTFCpd1	-0.123	-0.097	0.122	-0.043	-0.046	-0.177
RSTGpd1	-0.136	-0.117	-0.092	-0.158	-0.097	-0.132
RPGa3	0.089	0.063	0.183	0.034	0.078	-0.017
RPG3	-0.077	-0.077	0.057	-0.009	-0.065	-0.124
RPOC1	-0.216	-0.123	0.192	-0.114	-0.015	-0.173
RSGad	0.097	0.153	0.190	0.202	0.210	0.106
RCGpd2	-0.106	-0.136	0.093	0.078	-0.090	-0.160
RSTGpd2	-0.126	-0.125	0.083	-0.082	-0.089	-0.185
RPOC2	-0.075	-0.050	-0.078	-0.128	-0.120	-0.090
RITGpd2	-0.232	-0.230	-0.178	-0.225	-0.180	-0.227
RPGpd	-0.149	0.026	0.267	-	0.035	-0.116
RTFCpd2	-0.070	-0.025	-0.067	-0.076	-0.094	-0.206
RMTGtp1	-0.056	-0.009	0.107	-0.035	-0.023	-0.027
RPC1	-0.178	-0.117	0.004	0.092	-0.115	-0.181
RSPL1	-0.105	-0.206	0.054	0.062	0.000	-0.177
RPG4	0.012	-0.026	0.107	-0.024	-0.022	-0.151
RSGp	0.090	0.100	0.112	0.132	0.177	0.065
RSPL2	0.020	-0.104	0.114	-0.017	0.034	-0.123
RMTGtp2	-0.141	-0.203	-0.125	-0.129	-0.087	-0.139
RITGtp	-0.125	-0.198	-0.169	-0.157	-0.120	-0.167
RTOFC	-0.060	0.088	0.189	0.121	0.173	-0.116
RLG1	-0.122	-0.124	0.055	-	0.142	0.032
RAG	-0.039	-0.039	-0.088	-0.080	-0.148	-0.178
RLOCsd1	0.049	0.096	0.061	0.050	0.021	-0.083
RPC2	0.046	0.028	0.132	-0.018	0.045	0.000
RPC3	-0.060	0.002	0.001	-0.001	0.214	0.058
RLOCsd2	0.044	0.008	0.171	0.080	0.075	-0.023
RLOCsd3	0.071	0.076	0.203	0.126	0.123	0.015
RLOCid1	-0.135	-0.099	-0.142	-0.122	-0.141	-0.229
RLOCid2	-0.060	-0.099	0.084	-0.055	-0.056	-0.168
ROFG	0.054	0.003	0.106	0.017	0.007	-0.121
RIC	-0.109	-0.120	0.055	-0.100	0.100	-0.131
RLOCsd4	-0.144	-0.130	0.083	0.035	0.011	-0.135
RLOCsd5	0.050	0.034	0.148	0.093	0.056	0.037
RLOCsd6	-0.067	-0.010	0.074	0.062	0.004	-0.008
RLOCid3	-0.222	-0.197	-0.182	-0.281	-0.228	-0.293
RLG2	0.008	-0.021	0.119	0.114	0.112	-0.031
ROP1	0.014	-0.068	0.117	-0.028	0.014	-0.105
ROP2	-0.124	-0.106	-0.009	-0.190	-0.061	-0.107
ROP3	-0.092	-0.034	0.065	-0.095	-0.030	-0.047

Table SB6: Spearman partial correlation values computed for step-wise target control centrality and age, corrected for outdegree, for regions of interest (ROIs) in the right hemisphere. Missing values correspond to cases in which the step-wise target control centrality is zero for all subjects.

pval	Frontal	Limbic	Temporal	Sensorimotor	Parietal	Occipital
RFP1	0.612	0.063	0.180	0.705	0.028	0.934
RFP2	0.405	0.282	0.071	0.899	0.045	0.323
RFP3	0.194	0.027	0.001	0.334	0.039	0.308
RFP4	0.182	0.475	0.125	0.885	0.284	0.788
RFP5	0.104	0.057	0.019	0.930	0.259	0.745
RPG1	0.055	0.873	0.063	0.488	0.444	0.737
RFP6	0.025	0.137	0.398	0.315	0.722	0.214
RFP7	0.018	0.965	0.528	0.632	0.912	0.040
RFP8	0.557	0.449	0.866	0.236	0.591	0.068
RPG2	0.072	0.777	0.776	0.680	0.877	0.259
RFP9	0.255	0.470	0.586	0.217	0.991	0.730
RFP10	0.212	0.544	0.006	0.625	0.828	0.014
RMFG1	0.000	0.004	0.294	0.495	0.540	0.426
RMFG2	0.782	0.879	0.124	0.191	0.038	0.428
RIFGpt	0.100	0.185	0.982	0.546	0.817	0.131
RSFG1	0.006	0.008	0.783	0.086	0.544	0.109
RCGad1	0.003	0.011	0.558	0.307	0.156	0.089
RIC1	0.010	0.441	0.180	0.265	0.067	0.043
RFOC	0.040	0.871	0.466	0.522	0.336	0.933
RTP1	0.092	0.425	0.043	0.543	0.205	0.420
RTP2	0.017	0.858	0.907	0.610	0.408	0.518
RMFG3	0.031	0.034	0.315	0.054	0.732	0.353
RMFG4	0.006	0.021	0.995	0.174	0.755	0.349
RPGa1	0.348	0.799	0.337	0.132	0.292	0.421
RCGad2	0.027	0.020	0.660	0.217	0.221	0.044
RIC2	0.317	0.125	0.783	0.554	0.994	0.017
RTP3	0.001	0.071	0.162	0.114	0.501	0.049
RCOC	0.005	0.125	0.112	0.011	0.428	0.004
RMTGad	0.009	0.088	0.095	0.089	0.311	0.010
RPP	0.001	0.519	0.372	0.465	0.725	0.017
RSFG2	0.178	0.335	0.130	0.503	0.248	0.198
RJL	0.297	0.203	0.325	0.257	0.544	0.273
RTFCa	0.016	0.220	0.758	0.421	0.834	0.783
RPGa2	0.205	0.242	0.015	0.581	0.462	0.291
RPGp1	0.104	0.262	0.450	0.189	0.073	0.101
RPGp2	0.382	0.191	0.837	0.012	0.648	0.121
RHG	0.092	0.520	0.003	0.236	0.485	0.813
RCGpd1	0.154	0.012	0.866	0.053	0.074	0.044
RMTGpd	0.031	0.001	0.034	0.001	0.005	0.007
RITGpd1	0.611	0.690	0.109	0.782	0.271	0.545
RTFCpd1	0.111	0.208	0.113	0.577	0.553	0.021
RSTGpd1	0.078	0.129	0.231	0.040	0.208	0.087
RPGa3	0.247	0.415	0.017	0.659	0.313	0.828
RPG3	0.321	0.318	0.460	0.907	0.401	0.107
RPOC1	0.005	0.110	0.012	0.137	0.847	0.024
RSGad	0.210	0.046	0.013	0.008	0.006	0.170
RCGpd2	0.169	0.076	0.229	0.310	0.245	0.038
RSTGpd2	0.102	0.105	0.282	0.291	0.250	0.016
RPOC2	0.329	0.516	0.314	0.096	0.120	0.242
RITGpd2	0.002	0.003	0.020	0.003	0.019	0.003
RPGpd	0.052	0.739	0.000	-	0.651	0.133
RTFCpd2	0.362	0.747	0.388	0.325	0.223	0.007
RMTGtp1	0.472	0.909	0.164	0.651	0.765	0.724
RPC1	0.020	0.129	0.960	0.231	0.137	0.018
RSPL1	0.172	0.007	0.481	0.418	0.997	0.021
RPG4	0.875	0.735	0.164	0.759	0.775	0.050
RSGp	0.245	0.196	0.146	0.087	0.021	0.397
RSPL2	0.799	0.176	0.140	0.827	0.660	0.110
RMTGtp2	0.067	0.008	0.106	0.093	0.258	0.071
RITGtp	0.104	0.010	0.028	0.040	0.118	0.030
RTOFC	0.439	0.256	0.014	0.116	0.024	0.133
RLG1	0.113	0.107	0.474	-	0.065	0.682
RAG	0.612	0.618	0.255	0.300	0.055	0.020
RLOCsd1	0.528	0.211	0.429	0.520	0.783	0.281
RPC2	0.548	0.713	0.087	0.818	0.557	0.996
RPC3	0.436	0.975	0.987	0.985	0.005	0.456
RLOCsd2	0.569	0.919	0.025	0.297	0.329	0.761
RLOCsd3	0.356	0.325	0.008	0.101	0.110	0.850
RLOCid1	0.080	0.199	0.066	0.112	0.066	0.003
RLOCid2	0.439	0.199	0.273	0.480	0.467	0.029
ROFG	0.486	0.966	0.170	0.830	0.933	0.115
RIC	0.155	0.120	0.480	0.194	0.193	0.088
RLOCsd4	0.061	0.091	0.281	0.652	0.886	0.080
RLOCsd5	0.515	0.661	0.055	0.228	0.471	0.636
RLOCsd6	0.383	0.896	0.338	0.421	0.961	0.917
RLOCid3	0.004	0.010	0.018	0.000	0.003	0.000
RLG2	0.916	0.791	0.122	0.137	0.145	0.691
ROP1	0.859	0.379	0.130	0.716	0.852	0.175
ROP2	0.108	0.167	0.910	0.013	0.429	0.166
ROP3	0.232	0.663	0.397	0.217	0.701	0.543

Table SB7: Spearman partial correlation p-values computed for step-wise target control centrality and age, corrected for outdegree, for regions of interest (ROIs) in the right hemisphere. Missing values correspond to cases in which the step-wise target control centrality is zero for all subjects.

Bibliography

[1] F. Abdelnour, M. Dayan, O. Devinsky, T. Thesen, and A. Raj. Functional brain connectivity is predictable from anatomic network's Laplacian eigenstructure. *NeuroImage*, 172:728–739, May 2018.

[2] R. Agarwala, T. Barrett, J. Beck, D.A. Benson, C. Bollin, E. Bolton, D. Bourexis, J.R. Brister, S.H. Bryant, K. Canese, M. Cavanaugh, C. Charowhas, K. Clark, I. Dondoshansky, M. Feolo, L. Fitzpatrick, K. Funk, L.Y. Geer, V. Gorelenkov, A. Graeff, W. Hlavina, B. Holmes, M. Johnson, B. Kattman, V. Khotomlianski, A. Kimchi, M. Kimelman, M. Kimura, P. Kitts, W. Klimke, A. Kotliarov, S. Krasnov, A. Kuznetsov, M.J. Landrum, D. Landsman, S. Lathrop, J.M. Lee, C. Leubsdorf, Z. Lu, T.L. Madden, A. Marchler-Bauer, A. Malheiro, P. Meric, I. Karsch-Mizrachi, A. Mnev, T. Murphy, R. Orris, J. Ostell, C. O'Sullivan, V. Palanigobu, A.R. Panchenko, L. Phan, B. Pierov, K.D. Pruitt, K. Rodarmer, E.W. Sayers, V. Schneider, C.L. Schoch, G.D. Schuler, S.T. Sherry, K. Siyan, A. Soboleva, V. Soussov, G. Starchenko, T.A. Tatusova, F. Thibaud-Nissen, K. Todorov, B.W. Trawick, D. Vakatov, M. Ward, E. Yaschenko, A. Zasypkin, and K. Zbicz. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, 46(D1):D8–D13, January 2018.

[3] L. Airas. Hormonal and gender-related immune changes in multiple sclerosis. *Acta Neurologica Scandinavica*, 132(S199):62–70, 2015.

[4] S. Andrews. FastQC: A Quality Control Tool for High Throughput Sequence Data, 2010.

[5] M. Ansart, S. Epelbaum, G. Bassignana, A. Bône, S. Bottani, T. Cattai, R. Couronné, J. Faouzi, I. Koval, M. Louis, E. Thibeau-Sutre, J. Wen, A. Wild, N. Burgos, D. Dormont, O. Colliot, and S. Durrleman. Predicting the Progression of Mild Cognitive Impairment Using Machine Learning: A Systematic, Quantitative and Critical Review. *Medical Image Analysis*, page 101848, October 2020.

[6] A. Avena-Koenigsberger, B. Misic, and O. Sporns. Communication dynamics in complex brain networks. *Nature Reviews Neuroscience*, 19(1):17–33, January 2018.

[7] A.L. Barabási and R. Albert. Emergence of Scaling in Random Networks. *Science*, 286(5439):509–512, October 1999.

[8] A.L. Barabási and Z.N. Oltvai. Network biology: Understanding the cell's functional organization. *Nature Reviews Genetics*, 5(2):101–113, February 2004.

[9] A. Barrat, M. Barthélemy, and A. Vespignani. *Dynamical Processes on Complex Networks*. Cambridge University Press, Cambridge, 2008.

[10] S. Bartolucci, F. Caccioli, and P. Vivo. A percolation model for the emergence of the Bitcoin Lightning Network. *Scientific Reports*, 10(1):4488, March 2020.

[11] B. Barzel and A.L. Barabási. Universality in network dynamics. *Nature Physics*, 9(10):673–681, October 2013.

[12] G. Bassignana, J. Fransson, V. Henry, O. Colliot, V. Zujovic, and F.D.V. Fallani. Step-wise target controllability identifies dysregulated pathways of macrophage networks in multiple sclerosis. *arXiv:2003.08913 [q-bio]*, March 2020.

[13] D.R. Bentley, S. Balasubramanian, H.P. Swerdlow, G.P. Smith, J. Milton, C.G. Brown, K.P. Hall, D.J. Evers, C.L. Barnes, H.R. Bignell, J.M. Boutell, J. Bryant, R.J. Carter, R. Keira Cheetham, A.J. Cox, D.J. Ellis, M.R. Flatbush, N.A. Gormley, S.J. Humphray, L.J. Irving, M.S. Karbelashvili, S.M. Kirk, H. Li, X. Liu, K.S. Maisinger, L.J. Murray, B. Obradovic, T. Ost, M.L. Parkinson, M.R. Pratt, I.M.J. Rasolonjatovo, M.T. Reed, R. Rigatti, C. Rodighiero, M.T. Ross, A. Sabot, S.V. Sankar, A. Scally, G.P. Schroth, M.E. Smith, V.P. Smith, A. Spiridou, P.E. Torrance, S.S. Tzonev, E.H. Vermaas, K. Walter, X. Wu, L. Zhang, M.D. Alam, C. Anastasi, I.C. Aniebo, D.M.D. Bailey, I.R. Bancarz, S. Banerjee, S.G. Barbour, P.A. Baybayan, V.A. Benoit, K.F. Benson, C. Bevis, P.J. Black, A. Boodhun, J.S. Brennan, J.A. Bridgham, R.C. Brown, A.A. Brown, D.H. Buermann, A.A. Bundu, J.C. Burrows, N.P. Carter, N. Castillo, M. Chiara E. Catenazzi, S. Chang, R. Neil Cooley, N.R. Crake, O.O. Dada, K.D. Diakoumakos, B. Dominguez-Fernandez, D.J. Earnshaw, U.C. Egbujor, D.W. Elmore, S.S. Etchin, M.R. Ewan, M. Fedurco, L.J. Fraser, K.V. Fuentes Fajardo, W. Scott Furey, D. George, K.J. Gietzen, C.P. Goddard, G.S. Golda, P.A. Granieri, D.E. Green, D.L. Gustafson,

N.F. Hansen, K. Harnish, C.D. Haudenschild, N.I. Heyer, M.M. Hims, J.T. Ho, A.M. Horgan, K. Hoschler, S. Hurwitz, D.V. Ivanov, M.Q. Johnson, T. James, T.A. Huw Jones, G.D. Kang, T.H. Kerelska, A.D. Kersey, I. Khrebtukova, A.P. Kindwall, Z. Kingsbury, P.I. Kokko-Gonzales, A. Kumar, M.A. Laurent, C.T. Lawley, S.E. Lee, X. Lee, A.K. Liao, J.A. Loch, M. Lok, S. Luo, R.M. Mammen, J.W. Martin, P.G. McCauley, P. McNitt, P. Mehta, K.W. Moon, J.W. Mullens, T. Newington, Z. Ning, B. Ling Ng, S.M. Novo, M.J. O'Neill, M.A. Osborne, A. Osnowski, O. Ostadan, L.L. Paraschos, L. Pickering, A.C. Pike, A.C. Pike, D. Chris Pinkard, D.P. Pliskin, J. Podhasky, V.J. Quijano, C. Raczy, V.H. Rae, S.R. Rawlings, A. Chiva Rodriguez, P.M. Roe, J. Rogers, M.C. Rogert Bacigalupo, N. Romanov, A. Romieu, R.K. Roth, N.J. Rourke, S.T. Ruediger, E. Rusman, R.M. Sanches-Kuiper, M.R. Schenker, J.M. Seoane, R.J. Shaw, M.K. Shiver, S.W. Short, N.L. Sizto, J.P. Sluis, M.A. Smith, J. Ernest Sohna Sohna, E.J. Spence, K. Stevens, N. Sutton, L. Szajkowski, C.L. Tregidgo, G. Turcatti, S. vandeVondele, Y. Verhovsky, S.M. Virk, S. Wakelin, G.C. Walcott, J. Wang, G.J. Worsley, J. Yan, L. Yau, M. Zuerlein, J. Rogers, J.C. Mullikin, M.E. Hurles, N.J. McCooke, J.S. West, F.L. Oaks, P.L. Lundberg, D. Klenerman, R. Durbin, and A.J. Smith. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456 (7218):53–59, November 2008.

[14] R.F. Betzel, S. Gu, J.D. Medaglia, F. Pasqualetti, and D.S. Bassett. Optimally controlling the human connectome: The role of network topology. *Scientific Reports*, 6, July 2016.

[15] N. Biggs. *Algebraic Graph Theory*. Cambridge University Press, May 1974.

[16] M. Bikson, P. Grossman, C. Thomas, A.L. Zannou, J. Jiang, T. Adnan, A.P. Mourdoukoutas, G. Kronberg, D. Truong, P. Boggio, A.R. Brunoni, L. Charvet, F. Fregni, B. Fritsch, B. Gillick, R.H. Hamilton, B.M. Hampstead, R. Jankord, A. Kirton, H. Knotkova, D. Liebetanz, A. Liu, C. Loo, M.A. Nitsche, J. Reis, J.D. Richardson, A. Rotenberg, P.E. Turkeltaub, and A.J. Woods. Safety of Transcranial Direct Current Stimulation: Evidence Based Update 2016. *Brain Stimulation*, 9(5):641–661, 2016 Sep-Oct.

[17] A. Bitsch, T. Kuhlmann, C.D. Costa, S. Bunkowski, T. Polak, and W. Brück. Tumour necrosis factor alpha mRNA expression in early multiple sclerosis lesions: Correlation with demyelinating activity and oligodendrocyte pathology. *Glia*, 29(4):366–375, 2000.

[18] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.U. Hwang. Complex

networks: Structure and dynamics. *Physics Reports*, 424(4):175–308, February 2006.

[19] P. Bonifazi, M. Goldin, M.A. Picardo, I. Jorquera, A. Cattani, G. Bianconi, A. Represa, Y. Ben-Ari, and R. Cossart. GABAergic Hub Neurons Orchestrate Synchrony in Developing Hippocampal Networks. *Science*, 326(5958):1419–1424, December 2009.

[20] M. Botvinick and T. Braver. Motivation and Cognitive Control: From Behavior to Neural Mechanism. *Annual Review of Psychology*, 66(1):83–113, 2015.

[21] J.A. Brown and J.D. Van Horn. Connected brains and minds—The UMCD repository for brain connectivity matrices. *NeuroImage*, 124:1238–1241, January 2016.

[22] E. Bullmore and O. Sporns. Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews. Neuroscience*, 10 (3):186–198, March 2009.

[23] E.T. Bullmore and D.S. Bassett. Brain Graphs: Graphical Models of the Human Brain Connectome. *Annual Review of Clinical Psychology*, 7(1):113–140, March 2011.

[24] M.K. Camlibel. Popov–Belevitch–Hautus type controllability tests for linear complementarity systems. *Systems & Control Letters*, 56(5):381–387, May 2007.

[25] C.T. Chen. *Linear System Theory and Design, International 3rd. Edition.* Oxford University Press, New York, 3rd revised international ed edition edition, February 2009.

[26] G. Chen. Pinning control and synchronization on complex dynamical networks. *International Journal of Control, Automation and Systems*, 12(2):221–230, April 2014.

[27] L. Cheng, Z. Wu, Y. Fu, F. Miao, J. Sun, and S. Tong. Reorganization of functional brain networks during the recovery of stroke: A functional MRI study. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4132–4135, August 2012.

[28] J.K. Choi, U. Yu, O.J. Yoo, and S. Kim. Differential coexpression analysis using microarray data and its application to human cancer. *Bioinformatics*, 21(24):4348–4355, December 2005.

[29] F. Chu, M. Shi, C. Zheng, D. Shen, J. Zhu, X. Zheng, and L. Cui. The roles of macrophages and microglia in multiple sclerosis and experimental autoimmune encephalomyelitis. *Journal of Neuroimmunology*, 318:1–7, May 2018.

[30] D. Colombi, C. Poletto, E. Nakouné, H. Bourhy, and V. Colizza. Long-range movements coupled with heterogeneous incubation period sustain dog rabies at the national scale in Africa. *PLOS Neglected Tropical Diseases*, 14(5): e0008317, May 2020.

[31] C. Commault, J.M. Dion, D.H. Trinh, and T.H. Do. Sensor classification for the fault detection and isolation, a structural approach. *International Journal of Adaptive Control and Signal Processing*, 25(1):1–17, 2011.

[32] C. Commault and J. van der Woude. A Classification of Nodes for Structural Controllability. *IEEE Transactions on Automatic Control*, pages 1–1, 2016.

[33] C. Commault, J.M. Dion, and J.W. van der Woude. Characterization of generic properties of linear structured systems for efficient computations. *Kybernetika*, 38(5):[503]–520, 2002.

[34] C. Commault, J.M. Dion, and D.H. Trinh. Observability Preservation Under Sensor Failure. *IEEE Transactions on Automatic Control*, 53(6):1554–1559, July 2008.

[35] C. Commault, J. Van der Woude, and P. Frasca. Functional target controllability of networks: Structural properties and efficient algorithms. *IEEE Transactions on Network Science and Engineering*, pages 1–1, 2019.

[36] T.U. Consortium. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Research*, 47(D1):D506–D515, January 2019.

[37] I.D. Couzin, J. Krause, N.R. Franks, and S.A. Levin. Effective leadership and decision-making in animal groups on the move. *Nature*, 433(7025):513–516, February 2005.

[38] N.J. Cowan, E.J. Chastain, D.A. Vilhena, J.S. Freudenberg, and C.T. Bergstrom. Nodal Dynamics, Not Degree Distributions, Determine the Structural Controllability of Complex Networks. *PLOS ONE*, 7(6):e38398, June 2012.

[39] E. Czeizler, C. Gratie, W.K. Chiu, K. Kanhaiya, and I. Petre. Target Controllability of Linear Networks. In E. Bartocci, P. Lio, and N. Paoletti, editors, *Computational Methods in Systems Biology*, Lecture Notes in Computer Science, pages 67–81, Cham, 2016.

[40] E. Czeizler, K.C. Wu, C. Gratie, K. Kanhaiya, and I. Petre. Structural Target Controllability of Linear Networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 15(4):1217–1228, July 2018.

[41] M. Daianu, N. Jahanshad, T.M. Nir, C.R. Jack, M.W. Weiner, M.A. Bernstein, and P.M. Thompson. Rich club analysis in the Alzheimer's disease connectome reveals a relatively undisturbed structural core network. *Human Brain Mapping*, 36(8):3087–3103, 2015.

[42] J.S. Damoiseaux, C.F. Beckmann, E.J.S. Arigita, F. Barkhof, P. Scheltens, C.J. Stam, S.M. Smith, and S.a.R.B. Rombouts. Reduced resting-state brain activity in the "default network" in normal aging. *Cerebral Cortex*, 18(8): 1856–1864, August 2008.

[43] J. Davey, H.E. Thompson, G. Hallam, T. Karapanagiotidis, C. Murphy, I. De Caso, K. Krieger-Redwood, B.C. Bernhardt, J. Smallwood, and E. Jefferies. Exploring the role of the posterior middle temporal gyrus in semantic cognition: Integration of anterior temporal lobe with executive processes. *NeuroImage*, 137:165–177, August 2016.

[44] F. De Vico Fallani, J. Richiardi, M. Chavez, and S. Achard. Graph analysis of functional brain networks: Practical issues in translational neuroscience. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369 (1653):20130521, October 2014.

[45] D. Del Vecchio and R.M. Murray. *Biomolecular Feedback Systems*. Princeton University Press, Princeton, 1 edition edition, October 2014.

[46] D. Delpini, S. Battiston, G. Caldarelli, and M. Riccaboni. Portfolio diversification, differentiation and the robustness of holdings networks. *Applied Network Science*, 5(1):37, July 2020.

[47] J.M. Dion, C. Commault, and J. van der Woude. Generic properties and control of linear structured systems: A survey. *Automatica*, 39(7):1125–1144, July 2003.

[48] A. Dobin, C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T.R. Gingeras. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1):15–21, January 2013.

[49] V. Domínguez-García, V. Dakos, and S. Kéfi. Unveiling dimensions of stability in complex ecological networks. *Proceedings of the National Academy of Sciences*, 116(51):25714–25720, December 2019.

[50] R.C. Dorf and R.H. Bishop. *Modern Control Systems*. Prentice Hall, Upper Saddle River, NJ, 12 edition edition, July 2010.

[51] N.U.F. Dosenbach, D.A. Fair, F.M. Miezin, A.L. Cohen, K.K. Wenger, R.A.T. Dosenbach, M.D. Fox, A.Z. Snyder, J.L. Vincent, M.E. Raichle, B.L. Schlaggar, and S.E. Petersen. Distinct brain networks for adaptive and stable task control in humans. *Proceedings of the National Academy of Sciences*, 104(26): 11073–11078, June 2007.

[52] Y. Drier, M. Sheffer, and E. Domany. Pathway-based personalized analysis of cancer. *Proceedings of the National Academy of Sciences*, 110(16):6388–6393, April 2013.

[53] V.M. Eguíluz, D.R. Chialvo, G.A. Cecchi, M. Baliki, and A.V. Apkarian. Scale-Free Brain Functional Networks. *Physical Review Letters*, 94(1):018102, January 2005.

[54] P. Erdös and A. Rényi. On the evolution of random graphs. *Pub. Math. Inst. Huang. Acad. Sci.*, 5(1):17–60, 1960.

[55] F.D.V. Fallani, V. Latora, and M. Chavez. A Topological Criterion for Filtering Information in Complex Brain Networks. *PLOS Computational Biology*, 13(1):e1005305, January 2017.

[56] C.M. Filley and C.M. Cullum. Attention and vigilance functions in normal aging. *Applied Neuropsychology*, 1(1-2):29–32, November 1994.

[57] A. Fornito, A. Zalesky, and E.T. Bullmore. *Fundamentals of Brain Network Analysis*. Elsevier, AP, Academic Press is an imprint of Elsevier, Amsterdam Boston Heidelberg London, 2016.

[58] J. Gao, Y.Y. Liu, R.M. D'Souza, and A.L. Barabási. Target control of complex networks. *Nature Communications*, 5:5415, November 2014.

[59] A. Ghasemian, H. Hosseinmardi, and A. Clauset. Evaluating Overfit and Underfit in Models of Network Community Structure. *IEEE Transactions on Knowledge and Data Engineering*, pages 1–1, 2019.

[60] M. Gilbert, G. Pullano, F. Pinotti, E. Valdano, C. Poletto, P.Y. Boëlle, E. D'Ortenzio, Y. Yazdanpanah, S.P. Eholie, M. Altmann, B. Gutierrez, M.U.G. Kraemer, and V. Colizza. Preparedness and vulnerability of African countries against importations of COVID-19: A modelling study. *The Lancet*, 395(10227):871–877, March 2020.

[61] J. Gómez-Gardeñes and V. Latora. Entropy rate of diffusion processes on complex networks. *Physical Review E*, 78(6):065102, December 2008.

[62] J. Goñi, A. Avena-Koenigsberger, N. Velez de Mendizabal, M.P. van den Heuvel, R.F. Betzel, and O. Sporns. Exploring the Morphospace of Communication Efficiency in Complex Networks. *PLoS ONE*, 8(3):e58070, March 2013.

[63] J. Gonzalez-Astudillo, T. Cattai, G. Bassignana, M.C. Corsi, and F.D.V. Fallani. Network-based brain computer interfaces: Principles and applications. *arXiv:2006.13187 [q-bio]*, June 2020.

[64] C. Gray, L. Mitchell, and M. Roughan. Bayesian Inference of Network Structure From Information Cascades. *IEEE Transactions on Signal and Information Processing over Networks*, 6:371–381, 2020.

[65] S. Gu, F. Pasqualetti, M. Cieslak, Q.K. Telesford, A.B. Yu, A.E. Kahn, J.D. Medaglia, J.M. Vettel, M.B. Miller, S.T. Grafton, and D.S. Bassett. Controllability of structural brain networks. *Nature Communications*, 6:8414, October 2015.

[66] S. Gu, R.F. Betzel, M.G. Mattar, M. Cieslak, P.R. Delio, S.T. Grafton, F. Pasqualetti, and D.S. Bassett. Optimal trajectories of brain state transitions. *NeuroImage*, 148(Supplement C):305–317, March 2017.

[67] J. Guillon, M. Chavez, F. Battiston, Y. Attal, V. La Corte, M. Thiebaut de Schotten, B. Dubois, D. Schwartz, O. Colliot, and F. De Vico Fallani. Disrupted core-periphery structure of multimodal brain networks in Alzheimer's disease. *Network Neuroscience*, 3(2):635–652, January 2019.

[68] M. Hallett. Transcranial magnetic stimulation and the human brain. *Nature*, 406(6792):147–150, July 2000.

[69] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. Springer, New York, NY, 2nd edition edition, 2016.

[70] S.L. Hauser, J.R. Oksenberg, and S.E. Baranzini. Multiple Sclerosis. In *Rosenberg's Molecular and Genetic Basis of Neurological and Psychiatric Disease*, pages 1001–1014. Elsevier, 2015.

[71] M. Hautus. Stabilization controllability and observability of linear autonomous systems. *Indagationes Mathematicae (Proceedings)*, 73:448–455, 1970.

[72] T.F. Heatherton and D.D. Wagner. Cognitive neuroscience of self-regulation failure. *Trends in Cognitive Sciences*, 15(3):132–139, March 2011.

[73] L. Hedayatifar, R.A. Rigg, Y. Bar-Yam, and A.J. Morales. US social fragmentation at multiple scales. *Journal of The Royal Society Interface*, 16(159): 20190509, October 2019.

[74] V.J. Henry, F. Saïs, E. Marchadier, J. Dibie, A. Goelzer, and V. Fromion. BiPOm: Biological interlocked Process Ontology for metabolism. How to infer molecule knowledge from biological process? In *International Conference on Biomedical Ontology, ICBO 2017*, page np, Newcastle upon Tyne, United Kingdom, September 2017.

[75] C. Heuberger and S. Wagner. The number of maximum matchings in a tree. *Discrete Mathematics*, 311(21):2512–2542, November 2011.

[76] J.E. Hopcroft and R.M. Karp. A n5/2 algorithm for maximum matchings in bipartite. In *12th Annual Symposium on Switching and Automata Theory (Swat 1971)*, pages 122–125, October 1971.

[77] P.A. Iglesias and B.P. Ingalls, editors. *Control Theory and Systems Biology*. The MIT Press, Cambridge, Mass, 1st edition edition edition, September 2009.

[78] E.M. Izhikevich. *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting*. Computational Neuroscience. MIT Press, Cambridge, Mass, 2007.

[79] H. Jeong, S.P. Mason, A.L. Barabási, and Z.N. Oltvai. Lethality and centrality in protein networks. *Nature*, 411(6833):41–42, May 2001.

[80] T. Jia and A.L. Barabási. Control Capacity and A Random Sampling Method in Exploring Controllability of Complex Networks. *Scientific Reports*, 3(1): 2354, August 2013.

[81] T. Jia, Y.Y. Liu, E. Csóka, M. Pósfai, J.J. Slotine, and A.L. Barabási. Emergence of bimodality in controlling complex networks. *Nature Communications*, 4:2002, June 2013.

[82] R.E. Kalman. Mathematical Description of Linear Dynamical Systems. *Journal of the Society for Industrial and Applied Mathematics Series A Control*, 1 (2):152–192, January 1963.

[83] K. Kanhaiya, E. Czeizler, C. Gratie, and I. Petre. Controlling Directed Protein Interaction Networks in Cancer. *Scientific Reports*, 7(1):10327, September 2017.

[84] T.M. Karrer, J.Z. Kim, J. Stiso, A.E. Kahn, F. Pasqualetti, U. Habel, and D.S. Bassett. A practical guide to methodological considerations in the controllability of structural brain networks. *Journal of Neural Engineering*, 17(2): 026031, April 2020.

[85] S. Kauffman. The ensemble approach to understand genetic regulatory networks. *Physica A: Statistical Mechanics and its Applications*, 340(4):733–740, 2004.

[86] C. Kerepesi, B. Szalkai, B. Varga, and V. Grolmusz. How to Direct the Edges of the Connectomes: Dynamics of the Consensus Connectomes and the Development of the Connections in the Human Brain. *PLOS ONE*, 11(6):e0158680, June 2016.

[87] J.Z. Kim, J.M. Soffer, A.E. Kahn, J.M. Vettel, F. Pasqualetti, and D.S. Bassett. Role of graph architecture in controlling dynamical networks with applications to neural systems. *Nature Physics*, advance online publication, September 2017.

[88] I. Klickstein, A. Shirin, and F. Sorrentino. Energy scaling of targeted optimal control of complex networks. *Nature Communications*, 8(1):15145, April 2017.

[89] I.S. Klickstein and F. Sorrentino. Control Distance and Energy Scaling of Complex Networks. *IEEE Transactions on Network Science and Engineering*, pages 1–1, 2019.

[90] O.M. Koper, J. Kamińska, K. Sawicki, and H. Kemona. CXCL9, CXCL10, CXCL11, and their receptor (CXCR3) in neuroinflammation and neurodegeneration. *Advances in Clinical and Experimental Medicine: Official Organ Wroclaw Medical University*, 27(6):849–856, June 2018.

[91] S. Kosack, M. Coscia, E. Smith, K. Albrecht, A.L. Barabási, and R. Hausmann. Functional structures of US state governments. *Proceedings of the National Academy of Sciences*, 115(46):11748–11753, November 2018.

[92] T.R. Lezon, J.R. Banavar, M. Cieplak, A. Maritan, and N.V. Fedoroff. Using the principle of entropy maximization to infer genetic interaction networks from gene expression patterns. *Proceedings of the National Academy of Sciences of the United States of America*, 103(50):19033–19038, December 2006.

[93] A. Li, S.P. Cornelius, Y.Y. Liu, L. Wang, and A.L. Barabási. The fundamental advantages of temporal networks. *arXiv:1607.06168 [nlin]*, July 2016.

[94] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16):2078–2079, August 2009.

[95] J. Li, X. Chen, S. Pequito, G.J. Pappas, and V.M. Preciado. Structural Target Controllability of Undirected Networks. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 6656–6661, Miami Beach, FL, December 2018.

[96] P.Y. Li. Lecture notes for the course 'Advanced Control System Design', 2016.

[97] X. Li, X. Wang, and G. Chen. Pinning a complex dynamical network to its equilibrium. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 51(10):2074–2087, October 2004.

[98] C.T. Lin. Structural controllability. *IEEE Transactions on Automatic Control*, 19(3):201–208, June 1974.

[99] C. Liseron-Monfils, A. Olson, and D. Ware. NECorr, a Tool to Rank Gene Importance in Biological Processes using Molecular Networks and Transcriptome Data. *bioRxiv*, page 326868, May 2018.

[100] J. Liu, X. Guan, and X. Ma. Interferon regulatory factor 1 is an essential and direct transcriptional activator for interferon {gamma}-induced RANTES/CCl5 expression in macrophages. *The Journal of Biological Chemistry*, 280(26):24347–24355, July 2005.

[101] Y. Liu and G. Liu. Number of maximum matchings of bipartite graphs with positive surplus. *Discrete Mathematics*, 274(1):311–318, January 2004.

[102] Y.Y. Liu and A.L. Barabási. Control Principles of Complex Networks. *Reviews of Modern Physics*, 88(3), September 2016.

[103] Y.Y. Liu, J.J. Slotine, and A.L. Barabási. Controllability of complex networks. *Nature*, 473(7346):167–173, May 2011.

[104] Y.Y. Liu, J.J. Slotine, and A.L. Barabási. Control Centrality and Hierarchical Structure in Complex Networks. *PLOS ONE*, 7(9):e44459, September 2012.

[105] H. Lodish, A. Berk, S.L. Zipursky, P. Matsudaira, D. Baltimore, and J. Darnell. Gene Replacement and Transgenic Animals. *Molecular Cell Biology. 4th edition*, 2000.

[106] G. Lohmann, D.S. Margulies, A. Horstmann, B. Pleger, J. Lepsien, D. Goldhahn, H. Schloegl, M. Stumvoll, A. Villringer, and R. Turner. Eigenvector Centrality Mapping for Analyzing Connectivity Patterns in fMRI Data of the Human Brain. *PLOS ONE*, 5(4):e10232, April 2010.

[107] M.I. Love, W. Huber, and S. Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 2014.

[108] D.G. Luenberger. *Introduction to Dynamic Systems: Theory, Models, and Applications*. Wiley, New York, 1 edition edition, May 1979.

[109] J.B. Lugagne, S. Sosa Carrillo, M. Kirch, A. Köhler, G. Batt, and P. Hersen. Balancing a genetic toggle switch by real-time feedback control and periodic forcing. *Nature Communications*, 8, November 2017.

[110] S. Maslov and K. Sneppen. Specificity and Stability in Topology of Protein Networks. *Science*, 296(5569):910–913, May 2002.

[111] S.M. McCormick and N.M. Heller. Regulation of Macrophage, Dendritic Cell, and Microglial Phenotype and Function by the SOCS Proteins. *Frontiers in Immunology*, 6, 2015.

[112] J.D. Medaglia, F. Pasqualetti, R.H. Hamilton, S.L. Thompson-Schill, and D.S. Bassett. Brain and cognitive reserve: Translation via network control theory. *Neuroscience & Biobehavioral Reviews*, 75:53–64, April 2017.

[113] T. Menara, D. Bassett, and F. Pasqualetti. Structural Controllability of Symmetric Networks. *IEEE Transactions on Automatic Control*, pages 1–1, 2018.

[114] T. Menara, S. Gu, D.S. Bassett, and F. Pasqualetti. On Structural Controllability of Symmetric (Brain) Networks. *arXiv:1706.05120 [cs, math]*, June 2017.

[115] J. Menche, A. Sharma, M. Kitsak, S.D. Ghiassian, M. Vidal, J. Loscalzo, and A.L. Barabási. Disease networks. Uncovering disease-disease relationships through the incomplete interactome. *Science (New York, N.Y.)*, 347(6224): 1257601, February 2015.

[116] F. Menolascina, G. Fiore, E. Orabona, L.D. Stefano, M. Ferry, J. Hasty, M. di Bernardo, and D. di Bernardo. In-Vivo Real-Time Control of Protein Expression from Endogenous and Synthetic Gene Networks. *PLOS Computational Biology*, 10(5):e1003625, May 2014.

[117] D.M. Mosser and J.P. Edwards. Exploring the full spectrum of macrophage activation. *Nature Reviews Immunology*, 8(12):958–969, December 2008.

[118] B. Motik, B. Cuenca Grau, and U. Sattler. Structured objects in owl: Representation and reasoning. In *Proceeding of the 17th International Conference on World Wide Web - WWW '08*, page 555, Beijing, China, 2008.

[119] S.F. Muldoon, F. Pasqualetti, S. Gu, M. Cieslak, S.T. Grafton, J.M. Vettel, and D.S. Bassett. Stimulation-Based Control of Dynamic Brain Networks. *PLOS Computational Biology*, 12(9):e1005076, September 2016.

[120] C.G. Mullighan, S. Goorha, I. Radtke, C.B. Miller, E. Coustan-Smith, J.D. Dalton, K. Girtman, S. Mathew, J. Ma, S.B. Pounds, X. Su, C.H. Pui, M.V. Relling, W.E. Evans, S.A. Shurtleff, and J.R. Downing. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature*, 446(7137): 758–764, April 2007.

[121] K. Murota and S. Poljak. Note on a graph-theoretic criterion for structural output controllability. *IEEE Transactions on Automatic Control*, 35(8):939–942, August 1990.

[122] M.A. Musen. The Protégé Project: A Look Back and a Look Forward. *AI matters*, 1(4):4–12, June 2015.

[123] N.S. Narayanan, E.Y. Kimchi, and M. Laubach. Redundancy and Synergy of Neuronal Ensembles in Motor Cortex. *The Journal of Neuroscience*, 25(17): 4207–4216, 2005.

[124] M. Newman. *Networks: An Introduction*. Oxford University Press, Oxford, New York, March 2010.

[125] L. Noli, A. Capalbo, C. Ogilvie, Y. Khalaf, and D. Ilic. Discordant Growth of Monozygotic Twins Starts at the Blastocyst Stage: A Case Study. *Stem Cell Reports*, 5(6):946–953, December 2015.

[126] K.B. Nooner, S. Colcombe, R. Tobe, M. Mennes, M. Benedict, A. Moreno, L. Panek, S. Brown, S. Zavitz, Q. Li, S. Sikka, D. Gutman, S. Bangaru, R.T. Schlachter, S. Kamiel, A. Anwar, C. Hinz, M. Kaplan, A. Rachlin, S. Adelsberg, B. Cheung, R. Khanuja, C. Yan, C. Craddock, V. Calhoun, W. Courtney, M. King, D. Wood, C. Cox, C. Kelly, A. DiMartino, E. Petkova, P. Reiss, N. Duan, D. Thompsen, B. Biswal, B. Coffey, M. Hoptman, D.C. Javitt, N. Pomara, J. Sidtis, H. Koplewicz, F.X. Castellanos, B. Leventhal, and M. Milham.

The NKI-Rockland Sample: A Model for Accelerating the Pace of Discovery Science in Psychiatry. *Frontiers in Neuroscience*, 6, 2012.

[127] H. Oh, E.C. Mormino, C. Madison, A. Hayenga, A. Smiljic, and W.J. Jagust. β-Amyloid affects frontal and posterior brain networks in normal aging. *NeuroImage*, 54(3):1887–1895, February 2011.

[128] N. Oliver, B. Lepri, H. Sterly, R. Lambiotte, S. Deletaille, M.D. Nadai, E. Letouzé, A.A. Salah, R. Benjamins, C. Cattuto, V. Colizza, N. de Cordes, S.P. Fraiberger, T. Koebe, S. Lehmann, J. Murillo, A. Pentland, P.N. Pham, F. Pivetta, J. Saramäki, S.V. Scarpino, M. Tizzoni, S. Verhulst, and P. Vinck. Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. *Science Advances*, 6(23):eabc0764, June 2020.

[129] G. Orphanides and D. Reinberg. A unified theory of gene expression. *Cell*, 108(4):439–451, February 2002.

[130] F. Pasqualetti, S. Zampieri, and F. Bullo. Controllability Metrics, Limitations and Algorithms for Complex Networks. *IEEE Transactions on Control of Network Systems*, 1(1):40–52, March 2014.

[131] R. Pastor-Satorras and A. Vespignani. *Evolution and Structure of the Internet: A Statistical Physics Approach*. Cambridge University Press, Cambridge, reissue edition edition, July 2007.

[132] S. Pilosof, M.A. Porter, M. Pascual, and S. Kéfi. The multilayer nature of ecological networks. *Nature Ecology & Evolution*, 1(4):1–9, March 2017.

[133] E.A. Power. Collective ritual and social support networks in rural South India. *Proceedings of the Royal Society B: Biological Sciences*, 285(1879):20180023, May 2018.

[134] C.L. Pu, W.J. Pei, and A. Michaelson. Robustness analysis of network controllability. *Physica A: Statistical Mechanics and its Applications*, 391(18): 4420–4425, September 2012.

[135] A. Raj, A. Kuceyeski, and M. Weiner. A Network Diffusion Model of Disease Progression in Dementia. *Neuron*, 73(6):1204–1215, March 2012.

[136] A. Ramadiah, F. Caccioli, and D. Fricke. Reconstructing and stress testing credit networks. *Journal of Economic Dynamics and Control*, 111:103817, February 2020.

[137] V. Ravindran, S. V., and G. Bagler. Identification of critical regulatory genes in cancer signaling network using controllability analysis. *Physica A: Statistical Mechanics and its Applications*, 474:134–143, May 2017.

[138] V. Ravindran, J.C. Nacher, T. Akutsu, M. Ishitsuka, A. Osadcenco, V. Sunitha, G. Bagler, J.M. Schwartz, and D.L. Robertson. Network controllability analysis of intracellular signalling reveals viruses are actively controlling molecular systems. *Scientific Reports*, 9(1):2066, February 2019.

[139] S. Raza, K.A. Robertson, P.A. Lacaze, D. Page, A.J. Enright, P. Ghazal, and T.C. Freeman. A logic-based diagram of signalling pathways central to macrophage activation. *BMC Systems Biology*, 2:36, April 2008.

[140] D.S. Reich, F. Mechler, and J.D. Victor. Independent and Redundant Information in Nearby Cortical Neurons. *Science*, 294(5551):2566–2568, December 2001.

[141] D.M. Rentz, J.J. Locascio, J.A. Becker, E.K. Moran, E. Eng, R.L. Buckner, R.A. Sperling, and K.A. Johnson. Cognition, reserve, and amyloid deposition in normal aging. *Annals of Neurology*, 67(3):353–364, 2010.

[142] H. Reyngoudt, T. Claeys, L. Vlerick, S. Verleden, M. Acou, K. Deblaere, Y. De Deene, K. Audenaert, I. Goethals, and E. Achten. Age-related differences in metabolites in the posterior cingulate cortex and hippocampus of normal ageing brain: A 1H-MRS study. *European Journal of Radiology*, 81 (3):e223–e231, March 2012.

[143] C. Robert, X. Lu, A. Law, T.C. Freeman, and D.A. Hume. Macrophages.com: An on-line community resource for innate immunity research. *Immunobiology*, 216(11):1203–1211, November 2011.

[144] W.J. Rugh and T. Kailath. *Linear System Theory, 2nd Edition.* Pearson, Upper Saddle River, NJ, 2nd edition edition, August 1995.

[145] J. Ruths and D. Ruths. Control Profiles of Complex Networks. *Science*, 343 (6177):1373–1376, March 2014.

[146] N. Sahni, S. Yi, M. Taipale, J.I. Fuxman Bass, J. Coulombe-Huntington, F. Yang, J. Peng, J. Weile, G.I. Karras, Y. Wang, I.A. Kovács, A. Kamburov, I. Krykbaeva, M.H. Lam, G. Tucker, V. Khurana, A. Sharma, Y.Y. Liu, N. Yachie, Q. Zhong, Y. Shen, A. Palagi, A. San-Miguel, C. Fan, D. Balcha, A. Dricot, D.M. Jordan, J.M. Walsh, A.A. Shah, X. Yang, A.K. Stoyanova, A. Leighton, M.A. Calderwood, Y. Jacob, M.E. Cusick, K. Salehi-Ashtiani,

L.J. Whitesell, S. Sunyaev, B. Berger, A.L. Barabási, B. Charloteaux, D.E. Hill, T. Hao, F.P. Roth, Y. Xia, A.J.M. Walhout, S. Lindquist, and M. Vidal. Widespread macromolecular interaction perturbations in human genetic disorders. *Cell*, 161(3):647–660, April 2015.

[147] R. Sala-Llonch, D. Bartrés-Faz, and C. Junqué. Reorganization of brain networks in aging: A review of functional connectivity studies. *Frontiers in Psychology*, 6, 2015.

[148] F. Sambataro, V.P. Murty, J.H. Callicott, H.Y. Tan, S. Das, D.R. Weinberger, and V.S. Mattay. Age-related alterations in default mode network: Impact on working memory performance. *Neurobiology of Aging*, 31(5):839–852, May 2010.

[149] P. Sanz Leon, S.A. Knock, M.M. Woodman, L. Domide, J. Mersmann, A.R. McIntosh, and V. Jirsa. The Virtual Brain: A simulator of primate brain network dynamics. *Frontiers in Neuroinformatics*, 7, 2013.

[150] C. Seguin, A. Razi, and A. Zalesky. Inferring neural signalling directionality from undirected structural connectomes. *Nature Communications*, 10(1):1–13, September 2019.

[151] B. Sen, S.H. Chu, and K.K. Parhi. Ranking Regions, Edges and Classifying Tasks in Functional Brain Graphs by Sub-Graph Entropy. *Scientific Reports*, 9(1):1–20, May 2019.

[152] R. Shields and J. Pearson. Structural controllability of multiinput linear systems. *IEEE Transactions on Automatic Control*, 21(2):203–212, April 1976.

[153] J.J. Slotine and W. Li. *Applied Nonlinear Control*. Pearson, Englewood Cliffs, N.J, 1991.

[154] K. So, K. Ganguly, J. Jimenez, M.C. Gastpar, and J.M. Carmena. Redundant information encoding in primary motor cortex during natural and prosthetic motor control. *Journal of Computational Neuroscience*, 32(3):555–561, June 2012.

[155] L. Song, P. Langfelder, and S. Horvath. Comparison of co-expression measures: Mutual information, correlation, and model based indices. *BMC Bioinformatics*, 13(1):328, December 2012.

[156] E.D. Sontag. *Mathematical Control Theory: Deterministic Finite Dimensional Systems*. Texts in Applied Mathematics. Springer-Verlag, New York, second edition, 1998.

[157] C.J. Stam. Modern network science of neurological disorders. *Nature Reviews Neuroscience*, 15(10):683–695, October 2014.

[158] G. Stelzer, N. Rosen, I. Plaschkes, S. Zimmerman, M. Twik, S. Fishilevich, T.I. Stein, R. Nudel, I. Lieder, Y. Mazor, S. Kaplan, D. Dahary, D. Warshawsky, Y. Guan-Golan, A. Kohn, N. Rappaport, M. Safran, and D. Lancet. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Current Protocols in Bioinformatics*, 54(1):1.30.1–1.30.33, 2016.

[159] R. Steuer, J. Kurths, C.O. Daub, J. Weise, and J. Selbig. The mutual information: Detecting and evaluating dependencies between variables. *Bioinformatics*, 18(suppl_2):S231–S240, October 2002.

[160] O. Strauss, P.R. Dunbar, A. Bartlett, and A. Phillips. The immunophenotype of antigen presenting cells of the mononuclear phagocyte system in normal human liver – A systematic review. *Journal of Hepatology*, 62(2):458–468, February 2015.

[161] X. Sun, F. Hu, S. Wu, X. Qiu, P. Linel, and H. Wu. Controllability and stability analysis of large transcriptomic dynamic systems for host response to influenza infection in human. *Infectious Disease Modelling*, 1(1):52–70, October 2016.

[162] E. Tang and D.S. Bassett. Colloquium: Control of dynamics in brain networks. *Reviews of Modern Physics*, 90(3):031003, August 2018.

[163] E. Tang, C. Giusti, G.L. Baum, S. Gu, E. Pollock, A.E. Kahn, D.R. Roalf, T.M. Moore, K. Ruparel, R.C. Gur, R.E. Gur, T.D. Satterthwaite, and D.S. Bassett. Developmental increases in white matter network controllability support a growing diversity of brain dynamics. *Nature Communications*, 8(1): 1252, November 2017.

[164] I.W. Taylor, R. Linding, D. Warde-Farley, Y. Liu, C. Pesquita, D. Faria, S. Bull, T. Pawson, Q. Morris, and J.L. Wrana. Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nature Biotechnology*, 27(2):199–204, February 2009.

[165] R.C. Team. R: A Language and Environment for Statistical Computing, 2014.

[166] A.J. Thompson, B.L. Banwell, F. Barkhof, W.M. Carroll, T. Coetzee, G. Comi, J. Correale, F. Fazekas, M. Filippi, M.S. Freedman, K. Fujihara, S.L. Galetta, H.P. Hartung, L. Kappos, F.D. Lublin, R.A. Marrie, A.E. Miller, D.H. Miller, X. Montalban, E.M. Mowry, P.S. Sorensen, M. Tintoré, A.L. Traboulsee,

M. Trojano, B.M.J. Uitdehaag, S. Vukusic, E. Waubant, B.G. Weinshenker, S.C. Reingold, and J.A. Cohen. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *The Lancet. Neurology*, 17(2):162–173, February 2018.

[167] D. Tomasi and N.D. Volkow. Aging and functional brain networks. *Molecular Psychiatry*, 17(5):549–558, May 2012.

[168] T. Tong, Q. Gao, R. Guerrero, C. Ledig, L. Chen, and D. Rueckert. A novel grading biomarker for the prediction of conversion from mild cognitive impairment to Alzheimer's disease. *IEEE Transactions on Biomedical Engineering*, 64(1):155–165, 2017.

[169] L. Torres, K.S. Chan, and T. Eliassi-Rad. GLEE: Geometric Laplacian Eigenmap Embedding. *Journal of Complex Networks*, 8(2), April 2020.

[170] T. Uno. Algorithms for enumerating all perfect, maximum and maximal matchings in bipartite graphs. In H.W. Leong, H. Imai, and S. Jain, editors, *Algorithms and Computation*, Lecture Notes in Computer Science, pages 92–101, Berlin, Heidelberg, 1997.

[171] S. Uygun, C. Peng, M.D. Lehti-Shiu, R.L. Last, and S.H. Shiu. Utility and Limitations of Using Gene Expression Data to Identify Functional Associations. *PLoS Computational Biology*, 12(12), December 2016.

[172] H.J. van Waarde, M.K. Camlibel, and H.L. Trentelman. A Distance-Based Approach to Strong Target Control of Dynamical Networks. *IEEE Transactions on Automatic Control*, 62(12):6266–6277, December 2017.

[173] A. Vinayagam, T.E. Gibson, H.J. Lee, B. Yilmazel, C. Roesel, Y. Hu, Y. Kwon, A. Sharma, Y.Y. Liu, N. Perrimon, and A.L. Barabási. Controllability analysis of the directed human protein interaction network identifies disease genes and drug targets. *Proceedings of the National Academy of Sciences*, page 201603992, April 2016.

[174] S.G. Wagner. On the number of matchings of a tree. *European Journal of Combinatorics*, 28(4):1322–1330, May 2007.

[175] T. Wagner, A. Valero-Cabre, and A. Pascual-Leone. Noninvasive Human Brain Stimulation. *Annual review of biomedical engineering*, 9:527–65, February 2007.

[176] L.Z. Wang, R.Q. Su, Z.G. Huang, X. Wang, W.X. Wang, C. Grebogi, and Y.C. Lai. A geometrical approach to control and controllability of nonlinear dynamical networks. *Nature Communications*, 7:ncomms11323, April 2016.

[177] L. Wang, S. Wang, and W. Li. RSeQC: Quality control of RNA-seq experiments. *Bioinformatics*, 28(16):2184–2185, August 2012.

[178] X.F. Wang and G. Chen. Pinning control of scale-free dynamical networks. *Physica A: Statistical Mechanics and its Applications*, 310(3-4):521–531, July 2002.

[179] D.J. Watts and S.H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, June 1998.

[180] S.D. Williams and M.R. Patterson. Resistance and robustness of the global coral–symbiont network. *Ecology*, 101(5):e02990, 2020.

[181] D. Wilson and J. Moehlis. Clustered Desynchronization from High-Frequency Deep Brain Stimulation. *PLOS Computational Biology*, 11(12):e1004673, December 2015.

[182] H.M. Wilson. SOCS Proteins in Macrophage Polarization and Function. *Frontiers in Immunology*, 5, 2014.

[183] J.C. Worrell, J. Rumschlag, R.F. Betzel, O. Sporns, and B. Mišić. Optimized connectome architecture for sensory-motor integration. *Network Neuroscience*, 1(4):415–430, July 2017.

[184] L. Wu, Y. Shen, M. Li, and F.X. Wu*. Network Output Controllability-Based Method for Drug Target Identification. *IEEE Transactions on NanoBioscience*, 14(2):184–191, March 2015.

[185] S. Wuchty. Controllability in protein interaction networks. *Proceedings of the National Academy of Sciences of the United States of America*, 111(19): 7156–7160, May 2014.

[186] X. Xu, H. Yuan, and X. Lei. Activation and Connectivity within the Default Mode Network Contribute Independently to Future-Oriented Thought. *Scientific Reports*, 6(1):1–10, February 2016.

[187] G. Yan, P.E. Vértes, E.K. Towlson, Y.L. Chew, D.S. Walker, W.R. Schafer, and A.L. Barabási. Network control principles predict neuron function in the Caenorhabditis elegans connectome. *Nature*, advance online publication, October 2017.

[188] T. Yan, W. Wang, L. Yang, K. Chen, R. Chen, and Y. Han. Rich club disturbances of the human connectome from subjective cognitive decline to Alzheimer's disease. *Theranostics*, 8(12):3237–3255, 2018.

[189] K. Yang, H. Bai, Q. Ouyang, L. Lai, and C. Tang. Finding multiple target optimal intervention in disease-related molecular network. *Molecular Systems Biology*, 4:228, 2008.

[190] B.A. Yankner, T. Lu, and P. Loerch. The Aging Brain. *Annual Review of Pathology: Mechanisms of Disease*, 3(1):41–66, February 2008.

[191] J.G.T. Zañudo, G. Yang, and R. Albert. Structure-based control of complex networks with nonlinear dynamics. *Proceedings of the National Academy of Sciences*, June 2017.

[192] L. Zdeborová and M. Mézard. The number of matchings in random graphs. *Journal of Statistical Mechanics: Theory and Experiment*, 2006(05):P05003–P05003, May 2006.

[193] M. Zhang, N. Savill, D.S. Margulies, J. Smallwood, and E. Jefferies. Distinct individual differences in default mode network connectivity relate to off-task thought and text memory during reading. *Scientific Reports*, 9(1):16220, November 2019.

[194] X. Zhang, H. Wang, and T. Lv. Efficient target control of complex networks based on preferential matching. *PLoS ONE*, 12(4), April 2017.

[195] Y. Zhang, A. Garas, and I. Scholtes. Controllability of temporal networks: An analysis using higher-order networks. *arXiv:1701.06331 [physics]*, April 2017.

[196] C. Zhao, W.X. Wang, Y.Y. Liu, and J.J. Slotine. Intrinsic dynamics induce global symmetry in network controllability. *Scientific Reports*, 5(1):1–5, February 2015.

[197] J. Zhao, T.H. Yang, Y. Huang, and P. Holme. Ranking Candidate Disease Genes from Gene Expression and Protein Interaction: A Katz-Centrality Based Approach. *PLOS ONE*, 6(9):e24306, September 2011.

[198] Y. Zhu, L. Bai, P. Liang, S. Kang, H. Gao, and H. Yang. Disrupted brain connectivity networks in acute ischemic stroke patients. *Brain Imaging and Behavior*, 11(2):444–453, April 2017.