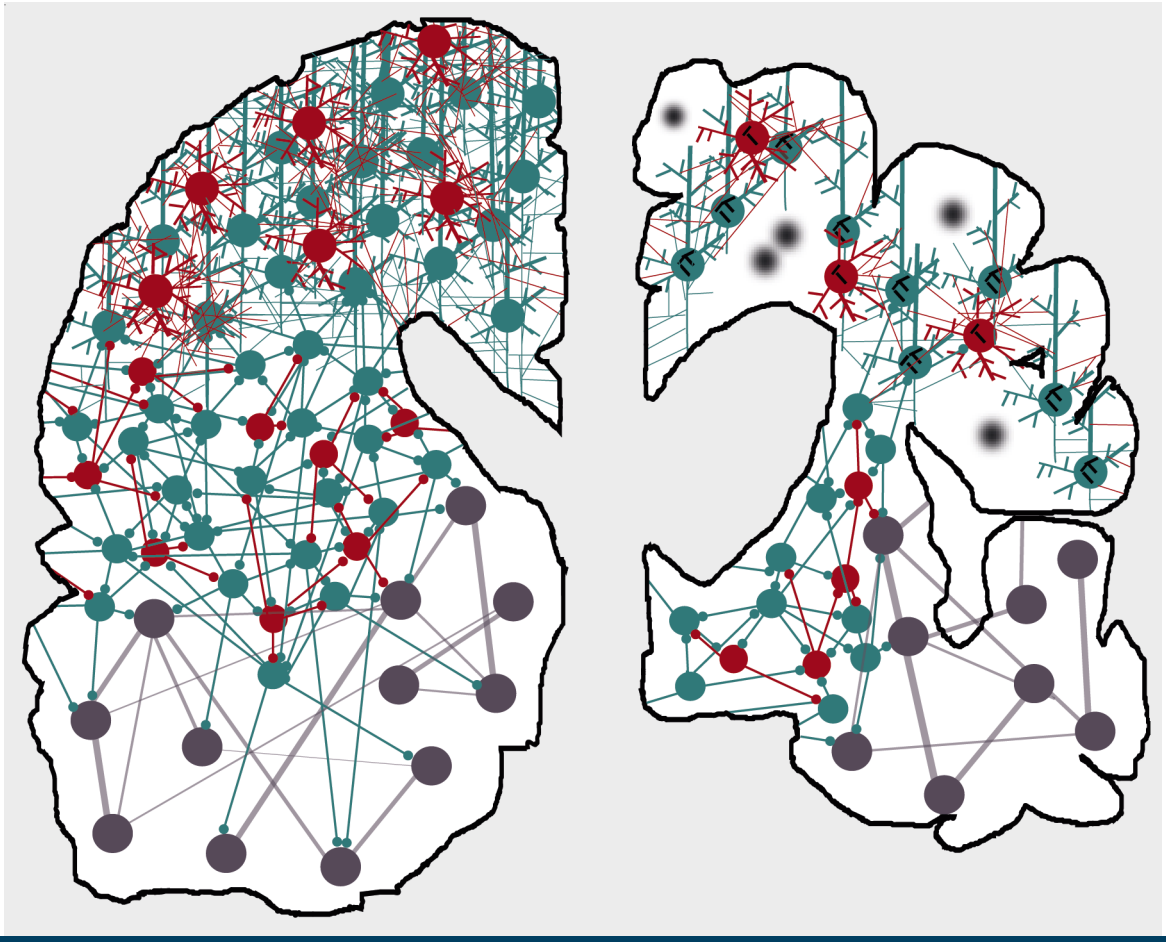

# Variability and compensation in Alzheimer‘s disease across different neuronal network scales

Claudia Bachmann

Schlüsseltechnologien / Key Technologies
Band / Volume 200
ISBN 978-3-95806-420-1

Mitglied der Helmholtz-Gemeinschaft

JÜLICH
Forschungszentrum

Schriften des Forschungszentrums Jülich
Reihe Schlüsseltechnologien / Key Technologies Band / Volume 200

Forschungszentrum Jülich GmbH
Institute of Neuroscience and Medicine
Computational and Systems Neuroscience (INM-6 / IAS-6)

# Variability and compensation in Alzheimer‘s disease across different neuronal network scales

Claudia Bachmann

Schriften des Forschungszentrums Jülich
Reihe Schlüsseltechnologien / Key Technologies Band / Volume 200

ISSN 1866-1807 ISBN 978-3-95806-420-1

Bibliografische Information der Deutschen Nationalbibliothek.
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der
Deutschen Nationalbibliografie; detaillierte Bibliografische Daten
sind im Internet über http://dnb.d-nb.de abrufbar.

| | |
|---|---|
| Herausgeber und Vertrieb: | Forschungszentrum Jülich GmbH<br>Zentralbibliothek, Verlag<br>52425 Jülich<br>Tel.: +49 2461 61-5368<br>Fax: +49 2461 61-6103<br>zb-publikation@fz-juelich.de<br>www.fz-juelich.de/zb |
| Umschlaggestaltung: | Grafische Medien, Forschungszentrum Jülich GmbH |
| Druck: | Grafische Medien, Forschungszentrum Jülich GmbH |
| Copyright: | Forschungszentrum Jülich 2019 |

Schriften des Forschungszentrums Jülich
Reihe Schlüsseltechnologien / Key Technologies, Band / Volume 200

D 82 (Diss. RWTH Aachen University, 2019)

ISSN 1866-1807
ISBN 978-3-95806-420-1

Vollständig frei verfügbar über das Publikationsportal des Forschungszentrums Jülich (JuSER)
unter www.fz-juelich.de/zb/openaccess.

# Abstract

Every human is unique and so is her diseases. This statement seems trivial but its consequences are far-reaching, especially for researchers and medical doctors trying to investigate and diagnose diseases. Some diseases progress in a stereotyped way, but many others show a variable phenotype. Especially diseases that interact with the intrinsic compensatory system are likely to feature manifold pathological changes. By observing individual, specific disease variables, in isolation, healthy and degenerated systems may be indistinguishable. It is mostly a combination of multiple variables that form the basis for disease understanding and diagnosis.

The pathology of Alzheimer's disease (AD) is associated with an inappropriate homeostatic compensation. The resulting complexity of this disease may be the reason for the two fundamental, unsolved challenges in AD. There is a lack of disease markers that can detect the disease onset in the preclinical phase itself. Moreover, there is no treatment that can effectively slow down the disease progression. The later might be a consequence of the poorly understood disease causes, which is aggravated by homeostatic interference. In this thesis the above stated difficulties in AD research are addressed in two different ways: The first part deals with the systematic investigation of a potential disease diagnosis tool. It is based on the structure of networks derived from functional magnetic resonance imaging (fMRI). The second part investigates the implication of AD and a particular type of homeostatic on the characteristics of small neuronal networks.

With respect to AD diagnosis, we construct brain graphs in which nodes represent brain areas and edges represent the functional connectivities. We then evaluate the resulting graph properties with respect to their diagnostic power, for three different health conditions: healthy, mild cognitive impaired and AD. We systematically examine which combinations of methods yield significant differences in the marginal distributions of the graph properties. The results are then evaluated with respect to consistency across different methods and predictability of diagnostic power. Crucial in these approaches is the definition of the diagnostic power, which is either based on a classification or on a probability measure. The latter can be directly combined with the results of other diagnostic tests, but requires the choice of an appropriate statistical model. Starting from first principles and approximations, we explain step-by-step how to construct such statistical models. In particular, we detail which models imply what assumptions on the data. In addition, we show how these statistical models can be evaluated and compared.

In the second part of this thesis, we use simulation to examine how the prominent synapse loss in AD (a network feature that best correlates with cognitive decline) affects computational performance of a simple recurrent network. We observe that deleting excitatory-excitatory synapses reduces the network's sensitivity to perturbations. It also increases generalization and reduces discrimination capability. Surprisingly, firing rate homeostasis based on an increase of the remaining excitatory-excitatory synapses, recovers performance for a wide range of lost connections. This phenomenon is examined further in an analytical model, substantiating the robustness of the results and providing more insight into underlying mechanisms.

# Überblick

Jeder Mensch ist einzigartig, genau wie seine Krankheiten. Diese Aussage hört sich trivial an, aber die resultierenden Konsequenzen sind weitreichend, vor allem für Forscher und Ärzte, die Krankheiten untersuchen und diagnostizieren wollen. Einige Krankheiten haben einen sehr stereotypischen Verlauf. Andere können dagegen einen sehr variablen Phänotyp aufweisen. Vor allem Krankheiten, die mit körpereigenen Kompensationsmechanismen interagieren, haben die Tendenz, verschiedene Symptome aufzuzeigen. Wenn nur eines dieser Symptome betrachtet wird, kann das kranke System unter Umständen nicht von einem gesunden unterschieden werden. Daher können diese Krankheiten nur verstanden und diagnostiziert werden, wenn viele Merkmale gleichzeitig betrachtet werden.

Die Alzheimer-Krankheit (AK) geht mit einer beeinträchtigten homöostatischen Kompensation einher. Die daraus resultierende Komplexität im Krankheitsbild trägt wahrscheinlich zu den zwei fundamentalen Herausforderungen in der AK-Forschung bei. Einerseits bedarf es mehr zuverlässiger Krankheitsmarker, die den Ausbruch der Krankheit schon in der vorklinischen Phase erkennen lassen. Andererseits gibt es noch keine Behandlung, die den Krankheitsverlauf verlangsamt. Letzteres resultiert daraus, dass die Krankheitsursache noch nicht genügend erforscht ist, was wiederum durch Kompensationsmechanismen erschwert wird. In dieser Arbeit gehen wir auf diese Herausforderungen auf zweierlei Arten ein. Im ersten Teil untersuchen wir ein potenzielles Diagnoseverfahren, welches auf der Untersuchung von Netzwerkstrukturen basiert, die aus funktionellen Magnetresonanztomografiedaten (fMRI) abgeleitet werden. Im zweiten Teil untersuchen wir die Auswirkungen der AK und einer bestimmten Art von Homöostase auf die Eigenschaften eines neuronalen Netzwerkes.

Im ersten Teil der Arbeit werden Graphen für die AK-Diagnose konstruiert, bei denen die Knoten Hirnareale und die Kanten funktionelle Verbindungen darstellen. Wir evaluieren die Grapheigenschaften bezüglich ihres Diagnosepotentials anhand von Patientendaten, und untersuchen systematisch, welche Methodenkombinationen signifikante Unterschiede zwischen den Marginalverteilungen der Grapheigenschaften hervorbringen. Die Ergebnisse werden in Bezug auf Robustheit, Praktikabilität und Diagnosevermögen evaluiert. Die Definition des Diagnosevermögens basiert entweder auf einem diskriminierenden oder generativen Modellierungsansatz. Bei letzterem wird ein statistisches Modell aus den Daten erstellt, das die Kombinierbarkeit mit anderen Diagnoseverfahren ermöglicht. Wir erläutern schrittweise, wie ein solches statistisches Modell konstruiert werden kann, und erklären, welches Modell welche Annahmen bezüglich der zugrunde liegenden Daten macht. Zusätzlich zeigen wir, wie die statistischen Modelle evaluiert und verglichen werden können.

Im zweiten Teil untersuchen wir die Auswirkungen von Synapsenverlust (ein prominentes Merkmal der AK) auf das rechnerische Leistungsvermögen eines einfachen neuronalen Netzwerkes. Wir beobachten, dass das Entfernen exzitatorischer Synapsen die Sensitivität des Netzwerkes reduziert. Gleichzeitig erhöht es die Generalisierungstendenz und verringert die Diskriminierungsfähigkeit. Wir zeigen, dass eine bestimmte Form von Homöostase, die die Feuerrate aufrechterhält, die Leistungsfähigkeit des Netzwerkes wiederherstellt. Mithilfe eines vereinfachten mathematischen Modells untermauern wir die Robustheit dieses Phänomens und gewinnen einen tieferen Einblick in die zugrunde liegenden Mechanismen.

# Declaration of contributions

This thesis is based on the collaboration of many researchers with varying degree of contributions. Especially the chapters 2-4 are a product of teamwork. In the following, I explicitly state the different contributions:

**Chapter 2:**

I designed the study, extracted graphs from fMRI data and performed all graph analysis. P.G.L. Porta Mana developed the statistical model. We both performed the statistical analysis of the data from the model. The work was carried out under the supervision of Abigail Morrison.

**Chapter 3:** I designed the study and constructed the graphs, applying several methods to cluster fMRI image data into nodes and define the edges between them. I calculated the graph properties and analyzed them with respect to significance, robustness and classification based on support vector machines. The statistical analysis was formulated and applied to the data in collaboration with P.G.L. Porta. Primary pre-processed fMRI data was provided by colleagues from INM-3 and Uniklink Köln, particularly K. Dillen and H.I.L. Jacobs. The work was carried out under the supervision of Abigail Morrison.

**Chapter 4:**

I developed the spiking neural network model for Alzheimer's disease, carried out the numerical simulation and analayzed the data. I constructed the analytical model and elaborated further theory-based conclusions in collaboration with T. Tetzlaff. The work was carried out under the supervision of Abigail Morrison.

# Acknowledgements

This work is a product of several years of work, in which many people supported me in various, sometimes quite difficult, situations. Here, I want to use the opportunity to thank all these people.

I thank Abigail Morrison very much for supervising me. During my PhD, some initially planned projects had to be revised due to unexpected obstacles (which I think is the normal case). Nevertheless, I have never felt left alone. I have experienced a friendly atmosphere with conversations and discussions on a peer level and decisions taking by mutual agreement.

I'm grateful to Björn Kampa for inspiring discussions and for evaluating this thesis.

I am also very thankful to Tom Tetzlaff for being my partner as well as for scientific discussions and collaborations. Although he is a physicist, he never loses the focus on biology. I really appreciate his profound understanding of scientific contexts and relationships and his great capability to explain difficult scientific context in a comprehensible, logical manner.

I thank PierGianLuca Porta Mana very much. I really enjoyed working with him. In particular I appreciate his idealistic attitude. He is smart and hard working and, at the same time, very helpful and patient to others.

Many thanks also to my colleagues and office mates. They have supported me a lot scientifically and have created a warm working environment! In particular, I want to thank the following people for giving constructive feedback on this thesis and proofreading it: Renato Duarte, Tom Tetzlaff, Nicole Voges, Jyotika Bahuguna, Alper Yegenoglu, Julius Korinth, Barna Zajzon, Philipp Weidel, Sandra Diaz, Rajalekshmi Deepu, Fahad Khalid and Charl Linssen.

I thank my family, friends and neighbors very much for supporting me in stressful situations with empathy and wise advices, cheerful conversations, distractions from work and childcare, when it was needed.

I am particularly thankful to my son Dragomir for being patient (with me being working), brave, open-minded and cheerful.

# Contents

# Chapter 1

# Introduction

The brain has to maintain its functionality and stable activity despite ongoing perturbations of different severity. These perturbations might be caused by external factors, e.g. displacement of the individual into a new environment or more intrinsically e.g. changes in protein expression patterns caused by different developmental stages. They might demand a change in behavior and, as a consequence, a change in brain activity, such that the overarching goal of an individual surviving and production of offspring is achieved. This challenges the brain to operate in a stable regime and, at the same time, adjust to new conditions. Such flexibility is achieved due to the rich repertoire of homeostatic regulatory mechanisms of the brain.

In this introduction, I will explain why this richness in homeostatic flexibility can cause problems when it comes to understanding disease mechanisms and making a disease diagnosis. Hereby, I will mainly focus on Alzheimer's disease (AD), because the core of this thesis deals with understanding and diagnosing AD in the face of a wide spectrum of homeostatic regulations occurring at different scales. In order to facilitate the understanding of the work presented in this thesis, short introductions into relevant topics are given in the following subsections Secs.1.1 -1.4.3. In the final subsection of this introduction (Sec.1.5), the relationship between homeostasis and disease challenges is explained. Subsequently, that reasoning forms a starting point for an outlook on the main topics and problems tackled in the remaining chapters of this work.

## 1.1 Alzheimer's disease - a disease with many facets

AD is a neurodegenerative disease that after years of progressive cognitive loss, leads to death. Unusually, there are two different manifestations of the disease: familiar and sporadic. Both have in common similar cognitive symptoms such as disorientation and memory loss, and morphologic abnormalities in brain structures like the abnormal accumulation of amyloid-$\beta$ (A$\beta$) plaques and tau tangles (see below). Familiar Alzheimer, the much rarer case, which only manifests in a person's thirties, forties, and fifties (for that reason also called early-onset-Alzheimer), is caused in almost all cases by a dominant acting mutation in one of the three

proteins: amyloid precursor protein (APP), presenilin-1 (PS1), or presenilin-2 (PS2). [1] In contrast, symptoms in sporadic (late-onset) AD develop in elderly individuals (65 and older) and the cause of this disease manifestation is still not known but widely debated. Since this disease type has a much higher prevalence (in the USA in 2018: $< 65$ years, $4\%$; $65 - 74$ years, $16\%$; $75 - 84$ years, $44\%$ ; $> 85$ years, $37\%$) [2] and still no treatment has been found that is capable of at least slowing down the disease progression, it causes high costs to the health care systems (13,080€ per patient per year in Germany; Reese et al., 2011) and thus research in sporadic AD is well financed.

Initially, the cognitive decline in AD is predominately characterized by loss of short-term memory. But, considering AD as a pure memory-deficit is an oversimplification. First, the notion of memory is very general and it is important to differentiate between the diverse subtypes such as episodic, semantic and working memory, which have different time courses during etiopathology. Second, Alzheimer patients, as well as individuals who are more likely to develop Alzheimer then cognitively unimpaired elderlies, show deficits in many other cognitive domains such as perceptual speed or visual-spatial processing (Bennett et al., 2002; Baldwin & Tomaszewski Farias, 2009).

This combination of different cognitive domains, which progressively become impaired, causes more and more difficulties to the AD patient in order to manage everyday life independently. Initially the patient might start repeating herself or displaces items more often. With further disease progression, the person might be less able to follow a conversation or appear confused due to non-logical jumps in her stream of thoughts. In moderate AD, executing more complex activities such as driving, cooking and medication management becomes difficult, if not even impossible. In the more advanced state, the patient needs support for basic activities such as eating, dressing or mobility. At this state, the person frequently loses the ability to recognize close friends and relatives. At the final state, other symptoms emerge such as motor and sensory problems, seizures and psychosis, eventually resulting in bed confinement.

The observed cognitive deterioration in AD is accompanied by structural and/or biochemical changes in the brain, some of which are very well visible even without microscopy including tissue atrophy, synapse and neuron loss, A$\beta$ plaques and tau tangles. More difficult to detect but not less important in understanding the disease is the deposition of A$\beta$ oligomers, synaptic plasticity impairment, increased microglial and astrocytic activation and altered oxidative stress response (see Fig.1.1). Among these pathological features, synapse loss correlates best with cognitive decline as post-mortem studies have substantiated (Terry et al., 1991; de Wilde et al., 2016). Already in the initial stages of the disease, a synapse loss of up to $15 - 25\%$ is observed in the lymbic system (comprising, amongst others, hippocampus, entorhinal cortex, amygdala, anterior nuclei of thalamus) and in the frontal cortex. In advanced AD this synapse loss increases up to $20 - 40\%$ and recruits gradually larger cortical brain areas (de Wilde et al., 2016). The loss of synapses and neurons is reflected in the observed shrinkage of brain tissue. Although the most reliable results are given by post-mortem autopsy, structural MRI can also give insight in how much brain tissue is lost.

[1] https://www.alzforum.org/early-onset-familial-ad/overview/what-early-onset-familial-alzheimer-disease-efad

[2] https://www.alz.org/documents_custom/2018-facts-and-figures.pdf

Depending on the particular brain region and on the disease severity, tissue atrophy up to $20\%$ is common and may even reach $40\%$ in the hippocampus of clinical AD patients. In a stereotyped view, brain atrophy first affects the medial temporal lobe (mainly entorhinal cortex and hippocampus) and then spreads along the temporal-parietal-frontal axis to the remaining cortical regions, whereas motor areas are generally spared. However, AD does not always proceed in a stereotyped fashion and two other atypical manifestations are known, out of which one is mainly characterized by neocortical atrophy and only small hippocampal tissue loss and the other mainly manifests in the limbic-system. (For review on brain atrophy measured via structural MRI see Pini et al. 2016.)

In terms of functional activity, e.g. as measured via functional MRI (fMRI, see Sec.1.4.1), observations across different studies do show both: increased and decreased BOLD signal for a couple of regions e.g. hippocampus and other regions within the default mode network (Vemuri et al., 2012; Damoiseaux, 2012). In contrast, studies based on FDG-PET, in which the glucose consumption in the brain is visualized, mainly report a decrease in glucose uptake (Marcus et al., 2014), which is interpreted as neuronal hypoactivity. However, two studies focusing on AD pre-stages also demonstrate a preceding hyperactivity (Poisnel et al., 2012; Herholz, 2010). The common consensus of these functional activation studies, which are based on different modalities, describes an initial hyperactivity of relevant brain regions followed by a long lasting hypoactivity that goes hand in hand with massive neuronal degeneration. It is still under debate whether the observed neuronal destruction results either from primary neurodestructive factors emerging as an initial AD trigger or are instead part of a compensatory reaction that may overreach a physiological state or may even prevent the neurons from even worse consequences.

Toxic A$\beta$ plaques and oligomers, apart from the cognitive symptoms, form the main link between familiar and sporadic AD. The amyloid precursor protein (APP) can be processed by two different pathways. In the non-amylogenic pathway the amyloid precursor protein (APP) is cleaved by an $\alpha-$secretase and subsequently by a $\gamma$- secretase leading to non-toxic cleavage products. In the amylogenic pathway, APP is cleaved by a $\beta-$secretase followed by a $\gamma$- secretase leading to toxic A$\beta$ aminoacids (see Fig.1.1). Depending on the exact length of the A$\beta$ molecules (40 or 42 aminoacids) these molecules are more or less hydrophobic and either conglomerate only slightly, forming soluble oligomers, or aggregate immensely, creating plaques. All mutations in familiar AD, which either affect the APP directly or the subunits of the $\gamma$-secretase (Presenilin 1 and 2) are said to increase the levels of A$\beta$ molecules in the traditional view. This is in line with observations made in sporadic AD, in which also increased A$\beta$ levels are found. However, more recent studies show that most of the mutations in the presenilin 1 gene which accounts for 95% of the familiar AD cases, rather lead to a loss-of-function of $\gamma$- secretase causing lower levels of A$\beta$ (Kelleher & Shen, 2015; Hunter & Brayne, 2017). This is only one example that highlights why A$\beta$ as an AD trigger is controversial and why other pathways related to $\gamma$- secretase activity (which does also cleave other membrane molecules) and APP functioning may also be important actors in disease progression.

Another key player in AD is the aggregated form of phosphorylated tau referred to as tau tangles. In the monomer configuration, its main physiological function lies in regulating microtubule dynamics, axonal transport and neurite outgrowth (Johnson & Stoothoff, 2004).

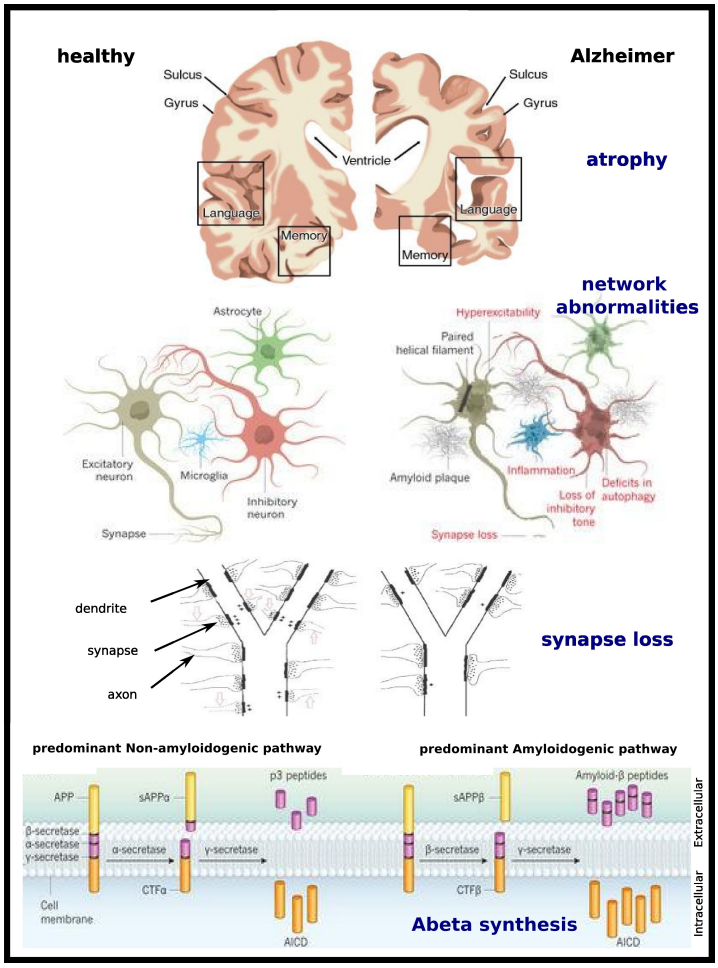


**Figure 1.1: Alzheimer pathology affects all scales in the brain.** This figure summarizes some prominent examples of pathogenic changes occurring in AD on different spacial scales by comparing healthy (left) and AD specific features (right): large areas in the brain suffer atrophy (upper panel); cell network abnormalities affect, on the one hand, neurons (tau tangles and neurodegeneration) and, on the other hand, glia cell mediated inflammation (second panel); on the neuronal level synapse loss is most prominent (third panel); on molecular level the cleavage of APP is shifted towards the amyloidogenic pathway in AD (last panel). Illustrations (second to fourth panel) with permission from Abuhassan et al. (2012) and Canter et al. (2016). The medical illustration (upper panel) is provided courtesy of Alzheimer's Disease Research, a BrightFocus Foundation program (*http://www.brightfocus.org/alzheimers)*.

The neurotoxic effect of its agglomeration is likely to be based on both: the lack of normal functioning tau and its toxic effect as tangle (Ballatore et al., 2007). Although it is largely accepted that tau -pathology is a consequence of A$\beta$ dysregulation, some studies report A$\beta$

independent pathways that trigger neurodegeneration in AD (Shen & Kelleher, 2007).

Despite the fact that a huge amount of money has been spent into AD research in the last years (the U.S. National Institute of Health will spend $1.414 billion to Alzheimer's disease and dementia research in 2019 [3]), and the knowledge about the disease is accumulating fast, the main fundamental problems still remain unsolved. So far, there exists no treatment which is able to at least slow down the disease progression. The most often prescribed acetylcholine esterase inhibitors and Memantine (an NMDA blocker) almost show no effect (Knight et al., 2018). The need to excuse this research failure, combined with the growing awareness of the manifold pathological changes in AD, which do not allow for reversal to a physiological state, have led to the second main challenge in the research focus: Identifying individuals that will develop AD as early as possible (predictive diagnosis).

Until today, the conclusive AD diagnosis is only made post-mortem. Since it is not possible to reliably diagnose AD in advanced disease stages in living individuals, it is even more challenging to find biomarkers that allow for a high sensitivity and selectivity with respect to an early AD diagnosis. In this regard, 'early' implies a diagnosis even at a time point when cognitive deficits are not even visible. The underlying reasoning is the following: The first morphological changes in AD are likely to be compensated by either homeostatic network mechanisms (compare Sec.1.3) and/or by the recruitment of cognitive resources that are usually, even if present, not used in the individual's daily life. The latter hypothesis originates from the observation that the process of cognitive decline is more slowly in individuals who experienced a higher level of education (Stern, 2012). The problem of an 'early' diagnosis is even intensified by the observation that some individuals with some cognitive complaints but who are still able to manage their daily life on their own (generally called mild cognitively impaired, MCI), do not develop AD. As a consequence, the early stages of AD are named and classified in various ways, e.g. 'probable ' (individuals with amnesic and nonamnesic disease indicators such as frontal dysfunction and language impairments), 'possible' (includes also atypical features), 'prodromal' (predementia stage with AD-type clinical amnesic phenotype and positive AD biomarkers) or 'preclinical' (no cognitive impairment but abnormal biomarkers relevant in AD ) (Chertkow et al., 2013). But how to identify individuals that will develop AD? And how can AD be distinguished from other types of dementia such as Lewy body dementia or frontotemporal dementia?

These questions can be addressed by intensifying research on AD biomarkers. In particular the following disease features, that are directly or indirectly accessible in the living individuals, are predominantly explored: A$\beta$ abnormalities measured in the CSF via lumbar puncture or through A$\beta$-PET, elevated levels of (phosphorylated) tau in the CSF, brain structure atrophy (mainly hippocampus) measured via volumetric MRI, cognitive tests such as Mini-Mental State Examination (MMSE) (for review on possible tests see Baldwin & Tomaszewski Farias, 2009), and synaptic dysfunction reflected in a decreased FDG-PET signal and in an altered functional activity measured via fMRI (Sperling et al., 2011). Only the first four are already used for AD diagnosis. Their sensitivity and specificity values rank from $80\%$ to $95\%$ (Humpel, 2011; Wattamwar & Mathuranath, 2010; Desikan et al., 2009) with biomarker- specific differences, e.g. A$\beta$ has a high sensitivity (all AD patients have

[3] https://www.aging.senate.gov/press-releases/nih-states-597-million-in-additional-fy-2019-funding-needed-for-alzheimers-research

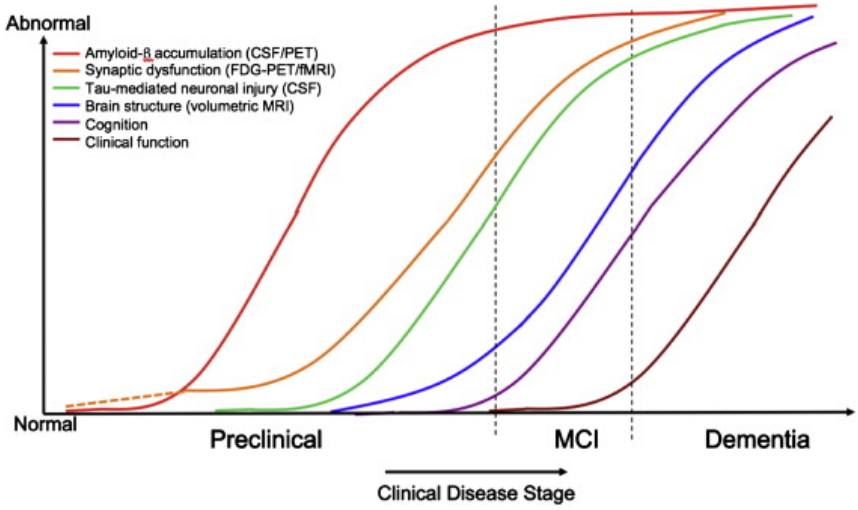


**Figure 1.2: Biomarker dynamics in AD.** Hypothetical change of AD biomarkers in the course of AD. Synaptic dysfunction might even be detectable before A$\beta$-abnormalities occur (dotted orange line). Abscissa: different stages of AD progression. Ordinate: biomarkers abnormalities. Figure reproduced with permission from Sperling et al. (2011)

increased A$\beta$ in the brain) but low specificity (healthy elderly individuals have also A$\beta$ plaques). Additionally, their dynamics in the course of the disease vary a lot, as shown in Fig.1.2. Most biomarkers change their profile only over a limited time window during the disease period ($A\beta$ concentrations change only in the beginning and reach their maximal abnormality already in the MCI state, cognitive tests can only detect MCI and latter stages). Only synaptic alterations appear to start already early on and reach their maximum only at very late stages. Unfortunately, FDG-PET and fMRI measures which are able to detect large scale synaptic alterations, are not currently used as diagnostic tools. This is probably due to the lack of standardized protocols allowing an accurate diagnosis. In particular, there is no common consensus on which particular information yields highest sensitivity and selectivity values across the entire disease spectrum and should therefore be used for diagnosis.

## 1.2 The concept of homeostasis in biology exemplified by neuronal network dynamics

In living systems all processes need to be highly regulated in order to keep the system in equilibrium, especially in the face of external perturbations and changes in the environment. This process is termed homeostasis and takes place on all scales, including microbiological structures, organelles, cells, groups of cells, tissues, organs and the entire body.

How this is achieved conceptually, can be explained by control theory, claiming the interaction of at least three necessary elements: a set point, a sensor and a negative feedback loop. A set point indicates the target values of the considered physiological process, which is referred to as regulated variable. It is not something that is defined through an intentional decision by the system itself, but rather something that arises as a product of evolutionary pressure and the system's intrinsic capability to account for it. Deviations from this set point are then detected by a sensor. This implies a direction specific change in the sensor's configuration, depending on whether the considered property overshoots or undershoots the set point. The change in the sensor's configuration triggers, via a controller, a negative feedback loop mediated through controlled variables, which counteract the deviation from the set point, bringing the system back to its equilibrium. All three elements (set point, sensor, negative feedback loop) can act simultaneously but eventually on different time scales (Cabanac, 2006; Kotas & Medzhitov, 2015).

The physiological range of the regulated variable can be either small, such that the system only allows for a single set point (e.g. arterial pressure of oxygen or blood calcium concentration in humans) or rather broad, such that multiple set points are possible (e.g. body weight, blood glucose, total body water) (see Fig.1.3, for details on terminology and more information see Kotas & Medzhitov 2015). If multiple set points are possible and the system stays for a long time in a regime that is far away from 'normal', e.g the person is highly overweighted, adverse long-term effects are the consequence. If a regulated variable is allowed to take on only a very small range of values and a disturbance is such, that the controlled variables cannot maintain the regulated variable at a stable level, parts of the system, if not the entire one, are likely to collapse. Sometimes, the need for compensation of a particular homeostatic systems interferes with the requirements of another one. In these cases, the overarching regulation system has to give preference to one of the systems, such that the consequences are least harmful. In such a case the overarching regulation system often gives preference to the regulated variables that operate only in a narrow range.

A regulated variable might be controlled by only a few or several controlled variables. These controlled variables, that refer to quantities, can also take small or large ranges of values, resulting in either small or broad distributions of their observed occurrence. In addition, only the combination of specific values of the controlled variables ensures the stability of a set point of the regulated variable. Thus, many combinations of controlled variables do not occur (see Fig.1.3).

An example of a regulated variable in a neuronal network is the firing rate of a neuron. In order to ensure that a network neither runs into permanent quiescence nor hyperactivity, the firing rates of the neurons have to stay in a certain range. In general, the firing rate of a neuron, among others, strongly depends on the neuron's excitability. In Van Welie et al. (2004) it is shown that this excitability can be regulated by at least three controlled variables: the amount of active AMPA receptors [4], NMDA receptors [5] and $I_{\mathrm{h}}$ [6] channels. Upregulating AMPA and NMDA receptors leads to an increase in excitation and hence to a higher firing

[4] $\alpha$-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor, ionotropic glutamate transmembrane receptor

[5] N-methyl-D-aspartate receptor, ionotropic glutamate transmembrane receptors

[6] hyperpolarization-activated cation current

rate. This can be counteracted by upregulating $I_h$ channels, which, in turn, lowers the firing rate again.

Another example of a regulated variable concerning neuronal activity is the spike bursting of cultured Purkinje cells [7] in mice (Swensen & Bean, 2005). Here, although some of the recorded neurons show almost identical spike bursting patterns in whole cells recordings, these patterns are generated based on different densities of Sodium and Calcium channels (the controlled variables) causing different current flows at different time points. (More examples are reviewed in Marder & Goaillard 2006).

These two examples show that certain processes of a neuron or neuronal network are under homeostatic control. Nevertheless this does not automatically imply that maintaining a target value of a certain regulated variable also retains the target values of another. This has been for example demonstrated in CA1 pyramidal neurons. Here, even if the firing rate of inter-neurons is fixed, differences in cellular or synaptic parameters alter network synchrony, which has been also found to underlie homeostatic regulation. (Santhakumar & Soltesz, 2004).

Coming back to the rather general concept of homeostasis, the phenomenon homeostasis can be also explained in a very different way. Thus, it can be considered as a product of dynamic self-organization. Accordingly, the dynamics of the considered $n$ properties underlying homeostasis are located in an $n$-dimensional attractor and small perturbations simply lead to a small displacement from the attractor. But, due to the attractor's force of attraction the dynamics will relax back to the center of the attractor after a certain relaxation time. The advantage of this approach is, that attractors can also have other dynamics then just fixed points (monotonic state of the observed properties) and hence can also describe periodic processes, such as oscillations. This extension of the understanding of homeostasis is incorporated in the concept of homeodynamics and is used for example in order to describe protein interactions (Lloyd et al., 2001).

## 1.3 Homeostasis in neuronal networks suffering Alzheimer's disease

The particular feature of neuronal cells is their ability to communicate via electrical signals, primarily via spikes. Therefore, not only processes that are typical for non-specialized cells, such as genome stability, proteostasis, energy supply, immune response and calcium signaling have to be regulated, but also network-wide spiking activity. In this section, I give a short overview about the mentioned homeostasis types and examples of their potential role in AD, mainly focusing on firing rate homeostasis as it is a main feature of neuronal activity (others are for example: firing synchrony or spiking regularity).

As explained in Sec.1.1, the hippocampus plays a major role in AD and various observations suggest that an impairment of its firing rate regulation plays a major role in AD progression.

[7] GABAergic neurons in the cerebellum.

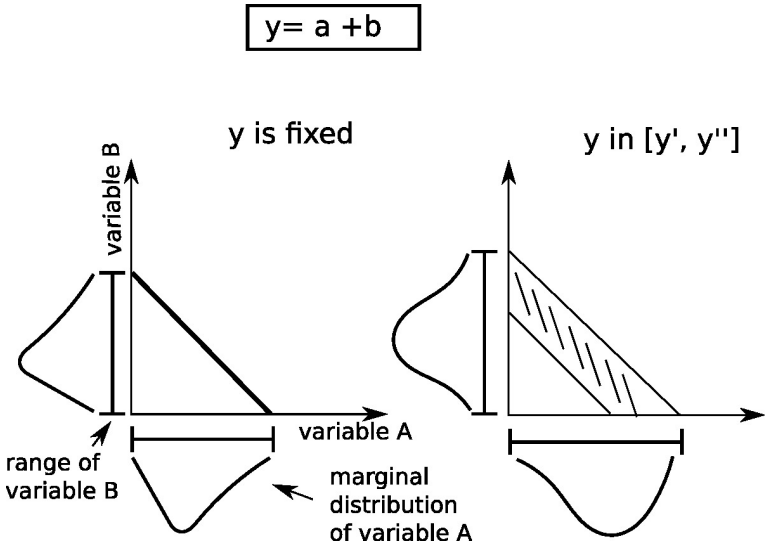


**Figure 1.3: Terminology of a homeostatic regulation system.** A regulated variable $Y$ manipulated by only two controlled variables $A$ and $B$ (with values $a$ and $b$), which for sake of simplicity interact linearly: $y = a + b$. Left: If $y$ is allowed to take on only a single value, only specific combinations of $a$ and $b$ are allowed (straight line). Although in this particular example the marginal distributions of $A$ and $B$ have to be the same, in general, the marginal distribution of one regulated variable does not entail information about another. Right: If the range of $Y$ is broader ($y \in [y\prime, y\prime\prime]$), fixing a particular value of $a$ still allows $b$ to take on a range of values and vice versa (hatched area). In this case, the marginal distributions of the two variables differ.

Evidence that firing rates are altered in the hippocampus of AD patients have been found in both human and mouse models. Although, at first glance studies have shown contradicting results and reported an increase as well as a decrease in the firing rates of hippocampal neurons, more and more evidence arose that both observations occur in AD, but hyperactivity seems to come first, followed by hypoactivity.

An extreme case of hyperactivity is an epileptic seizure, in which many neurons fire synchronously. Some standard EEG and MEG measurements have revealed a higher rate of epileptic-like discharges in AD patients compared with age-matched healthy controls (Mendez et al., 1994; Amatniek et al., 2006). Furthermore, recent intracranial recordings of two AD patients in the early disease phase have revealed silent hippocampal seizures during sleep, that were not measurable with normal EEG (Lam et al., 2017). Evidence for higher firing rates and increased neuronal excitability also comes from in vitro patch clamp recordings of hippocampal slices and in vivo investigations of different AD mouse types with A$\beta$ pathology (Hall et al., 2015; Šišková et al., 2014). These observations suggest that hyperactivity is a true hallmark of early disease stages. Decreased firing rates have been mainly reported for mouse models that mainly feature tau tangles and not A$\beta$ pathology (Menkes-Caspi et al., 2015; Fu et al., 2017).

As mentioned above, these results are not mutually exclusive, since increased and decreased

firing rates might occur at different time periods and involve different neuron types, in which A$\beta$ and tau derivatives might even be expressed and/or modified in different ways. More insight should be obtained by long-term recordings. A first step towards long-term recordings is done based on bulk $Ca^{2+}$-imaging in CA1 hippocampal familial AD mice, which express both A$\beta$- and tau- pathology. Here, very young mice (1- to 2-month old) only show hyperactive cells. With aging (6- to 7-month old) both hyperactive and hypoactive neurons have been found (Busche et al., 2012). This is in line with observations from PET and fMRI studies(see Sec.1.4.1), in which an increase followed by a decrease in glucose utilization and in the BOLD signal respectively has been shown (Dickerson et al., 2005; O'Brien et al., 2010; Herholz, 2010).

Firing rate alterations in neuronal networks can have several reasons, e.g. a change of the intrinsic properties of neurons, alterations in synaptic connectivity, modifications of the input to the network and/or if more than one stable attractor exist, the shift of the network dynamics to another attractor. All these scenarios highly influence each other, e.g. a shift of the neurons' properties will also affect the synaptic connectivity of the network in the long run. Most studies investigating underlying causes of AD focus on the first two points and their interactions. Here, special attention is given towards the damage of synapses, hence it highly correlates with the cognitive deficits observed in AD (see Sec.1.1). In the following paragraphs, I will give an example of what is dysregulated in AD for each of the aforementioned domains underlying homeostatic control (genome stability, proteostasis etc.) and the resulting effect on synapse functioning, which might result in firing rate alterations. (For an extensive overview see e.g. Frere & Slutsky 2017.)

Genome stability mainly refers to methylation and acetylation of the DNA strand, which, by regulating its accessibility for the transcription machinery, also controls gene expression. In AD patients, the occurrence of histone deacetylases is increased, reducing the acetylation level of histones and therefore hampering the expression of important transcription factors that regulate synaptic plasticity and learning (e.g. the activity-regulated cytoskeleton-associated protein $Arc$) (Gräff et al., 2012).

Proteostasis describes all mechanisms that influence the activity of a protein including protein synthesis, folding, transport and degradation. The most prominent misregulated proteins in AD are A$\beta$ and tau. In both cases, a bidirectional relationship with respect to synaptic activity and protein expression is known. In a simplified view, a high concentration of A$\beta$ oligomers inhibits synapse growing and impairs synaptic plasticity. The resulting reduced synaptic activity reduces A$\beta$ secretion (Kamenetz et al., 2003). Low concentrations of A$\beta$ might even enhance long term potentiation. The protein tau is normally located in axons. Increased synaptic activity (and probably the presence of A$\beta$) leads to a relocation to the post-synaptic terminals of the somatodentritic compartments and, as a consequence, to a loss of excitatory synapses (spines) and therefore a reduction in firing rate. As reviewed in Tampellini (2015), these processes are much more complex than it is described here and highly depend on the location and the exact structure of considered proteins.

Ionic $Ca^{2+}$ is one of the most important signaling molecules in neurons and its effects highly depend on its local concentration. It is well known for its role in regulating neurotransmitter release, signaling through voltage-gated calcium channels and ionotropic receptors, such as

NMDR. As mentioned above, its intracellular concentration is strictly regulated, otherwise its signaling function would be disturbed. This regulation also implies its removal from the intracellular matrix, which can be either done by the endoplasmic reticulum (ER), mitochondria or transport to extracellular space. In AD brains and familiar AD mice, an augmented $Ca^{2+}$ uptake from the ER to the mitochondria has been reported (Hedskog et al., 2013), which may induce mitochondrial dysfunction and stress responses, potentially leading to cell death (Rizzuto et al., 2012).

There is some evidence that disruption of energy homeostasis plays a crucial role in AD etiology. For example, patients with type 2 diabetes (Huang et al., 2014) and elevated glucose blood levels (even in the absence of diabetes; Crane et al., 2013), show an increased risk for developing AD-related dementia. Also a high-fat, low carbohydrate ketogenic diet, that shifts the brain metabolism towards ketone bodies as a major energy source and not glucose, has been found to improve cognitive performance (Reger et al., 2004). With respect to possible underlying molecular mechanisms (Frere & Slutsky, 2017), it is not clear how glucose metabolism triggers AD. More insight would be very important, since these are mechanisms that individuals, at least to a certain extent, can influence by adapting their lifestyle.

The activation of the immune system has been found to be both beneficial and damaging (for review see Frere & Slutsky, 2017). Microglia may directly modulate synapses and initiate synaptic pruning (Hong et al., 2016). The ablation of microglia, although not substantially affecting the formation of plaques, is thought to ameliorate the cognitive decline of familiar AD mice (Dagher et al., 2015; Grathwohl et al., 2009).

The above examples form only a small subset of already detected disrupted homeostatic mechanisms in AD. Since AD is accompanied by large morphological changes such as synapse and neuron loss, it is very difficult to decide whether the observed changes are a result of homeostatic dysregulation or whether they are part of a homeostatic response. For example, if the network is hyperactive, it is crucial for the overall survival of the network to delete synapses or even to kill some neurons. This might trigger an increased uptake of $Ca^{2+}$ in the mitochondria, which would lead to neuronal decay. Now, only considering $Ca^{2+}$ homeostasis would convey the (wrong) picture, that $Ca^{2+}$ homeostasis is disturbed in AD. Thus, it is important to disentangle the dependencies, by searching for initial events, testing whether a recovery of a specific homeostasis (via drugs) can counteract the cognitive decline or by identifying risk factors.

## 1.4 Methodological considerations

The investigation of AD and associated homeostatic systems, which is performed in this work, is based on methods that might not be familiar to the reader. In order to facilitate the understanding of the following chapters, I give a short introduction to the main methodical concepts, which are: functional magnetic resonance imaging (fMRI, Sec.1.4.1), graph theory (Sec.1.4.2), neuronal network simulations and how the performance of neuronal networks with respect to certain tasks can be quantified (Sec.1.4.3).

### 1.4.1 What does fMRI measure?

In general, magnetic resonance imaging (MRI) is able to identify different types of body tissue based on different relaxation times of protons after magnetization. A strong static magnetic field causes a longitudinal magnetization of the protons aligned to the static field. An additional high frequency alternating field causes a tilted rotation of the protons, transversal to the magnetic field, and induces an electric voltage in a receiving coil. After turning off the alternating magnetic field, the protons relax back to their original orientation. Depending on the nature of the respective atoms, the relaxation times differ. Thus, based on the relaxation time, different tissue types can be determined. Whereas in structural MRI, protons of hydrogen nuclei in water molecules are excited, functional MRI uses the paramagnetic properties of deoxygenated hemoglobin, which is magnetizable, in contrast to diamagnetic oxygenated hemoglobin.

In the 1890s, researchers discovered that the oxygenation level of hemoglobin, is related to neuronal activity. If neuronal activity is increased, the so called blood-oxygen-level dependent (BOLD) signal, increases around two seconds later, reaching its peak over four to six seconds before returning back to baseline level. The raising behavior of the BOLD signal is not trivial, because different processes with positive and negative signs sum up: the oxygen extraction fraction, which is the fraction of oxygen by an element of blood that moves from the blood vessels into the brain tissue through the capillary bed, rises due to elevated energy consumption and, as a consequence, the BOLD signal would decrease. But, at the same time, the cerebral blood flow increases, providing the brain with more oxygenated blood, and hence leads to a net increase of the BOLD signal. In addition, other effects such as changes in the cerebral blood volume, which, depending on the type of blood vessel (arterial or venous) can lower or magnify the BOLD signal (for review see Buxton, 2013).

Due to the multicomponent nature of the BOLD signal, changes in the fMRI signal, especially those obtained from comparing two different groups e.g. healthy or diseased, have to be interpreted with care. They could be caused by altered neuronal activity (which is the most common interpretation), but also by chronic alterations in the baseline oxygenation level of the blood or changes in the structure and functioning of blood vessels. In order to get more insight into what causes differences in the BOLD signal, calibrated fMRI can be applied. Here, the experimental protocol allows for a differentiation of the various underlying processes and thus a clarification of what exactly causes the observed disparities.

### 1.4.2 Introduction into graph theory

A graph is a mathematical structure that consists of nodes and edges, in which the edges indicate relations between the nodes. These relations can either be quantified yielding weighted graphs or be given as yes-or-no statements ('yes': an interaction between the two considered nodes is given, and 'no': it is not given) resulting in binary graphs. Furthermore, a direction of the interaction can be indicated, allowing for the differentiation between the influence that node 'A' has on node 'B' and vice versa (directed graph). But this information is not always accessible or it is not used because it complicates graph analysis and therefore

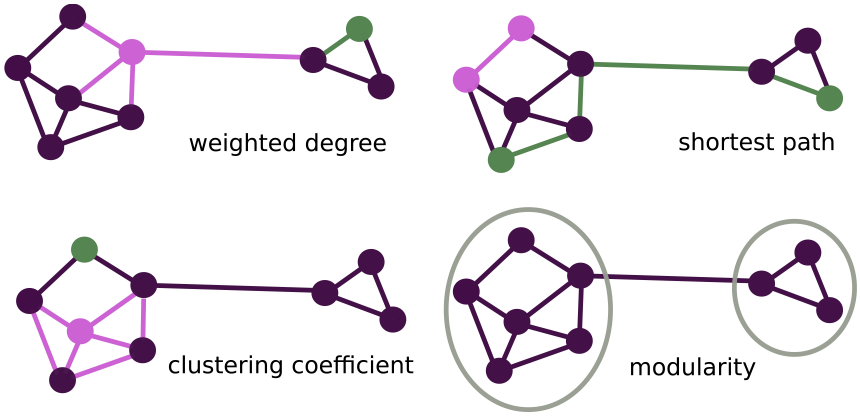


**Figure 1.4: Illustration of graph properties.** Based on a binary, undirected graph weighted degree, shortest path, clustering coefficient and modularity are illustrated. The weighted degree (upper left) measures the relative connection strength of a node with respect to all other nodes in the network. The pink node has a higher degree ($\frac{\text{number of existing edges}}{\text{number of possible edges}} = \frac{4}{8}$) then the green node ($\frac{1}{8}$). The clustering coefficient (lower left) reveals how dense direct node neighbors are connected among each other. The pink node has a clustering coefficient of $\frac{3}{4}$ because all four neighbors could form four connections, but only three exists. The green node has a clustering coefficient of $\frac{0}{1}$ The minimum shortest path (upper right) in a binary graph is one (pink connection between the two pink nodes), which accounts for a direct connection. The two green nodes share a very long shortest path of four, because at least four edges have to be crossed to reach the other node. The depicted graph can be divided into two modules (lower right, grey circles), with dense connectivity in each module but only a single edge connecting the two.

edges are defined reciprocally bringing about undirected graphs.

In general, graph theory analyzes the structure of graphs, which is mainly characterized by the exact arrangement and nature of graph edges. Depending on the specific scientific question, different graph properties might be investigated.

The importance of nodes in a network can differ depending on their position and connection strengths to other nodes. Thus, removal of a central lying nodes (which make many strong connections to other graph nodes) might cause a total disruption of the graph structure and a breakdown of the entire system, whereas deleting peripheral nodes might cause only marginal damage to the associated network.

In order to investigate the centrality of a node, different measures can be applied. The degree of a node sums up the strengths of its connections to the other graph nodes. The weighted degree results from normalizing the degree (see Fig.1.4) and yields values between zero and one, such that values closer to one indicate densely connected nodes. In protein-protein-interaction networks, in which the graph's edges describe the cooperation of different proteins, nodes with high degree have been found to be essential for cell survival (Jeong et al., 2001).

The centrality of a node can also be evaluated with respect to the strengths of its emanating paths to the other nodes in the graph. The so-called shortest path, which is the minimum of all possible sums taking over all edges of all possible paths connecting two nodes (see Fig.1.4), measures the potential ability of efficient information transmissions. Therefore, the normalized sum of all shortest paths emanating from a singe node, the so-called closeness centrality, indicates how central this node is with respect to information transmission. Striking short shortest paths are a general feature of metabolic networks describing biochemical reactions within a cell. Even with rising complexity of cellular processing going hand in hand with an augmented number of different substrates, the shorteness of these paths does not increase. This is due to a concomitant increase of the average number of reactions in which certain substrates participate (Jeong et al., 2000). In genome-based large scale metabolic networks, nodes of high closeness centrality indicate top central metabolites (Ma & Zeng, 2003).

The extent to which a node is connected locally is quantified by the clustering coefficient (see Fig.1.4). The local embedding not only includes how strong a node is connected to neighboring nodes but also how dense the neighboring nodes are connected among each other. Therefore, the clustering coefficient can be used, for example, to identify small and highly connected topologic modules in metabolic networks, forming functional entities fundamental for cellular organization (Ravasz et al., 2002).

Extending the analysis of graphs further towards graph meta-structures opens up new possibilities with respect to finding explanations for emergent phenomena that requires the analysis of the interaction between multiple nodes in an overarching fashion. The analysis of the meta-structure of a graph might reveal a partition of the graph into different modules in which nodes belonging to a module are highly connected but connections between modules are rare (see Fig.1.4). As a consequence, deleting intra-modular connections might have fatal consequences for the system, as opposed to erasing inter-modular edges. Thus, changes in the modular structure can lead to profound changes in the system's output. For example, a reduced modularity in functional brain graphs (see below) has been associated with the awareness of a visual target (Godwin et al., 2015).

With respect to neuroscience, graph theory is applied to brain networks derived from various spacial scales. Thus, graph nodes may correspond to brain-specific molecular structures and proteins (see e.g. Godwin et al., 2015), single neurons, small neuronal populations or entire brain areas (Godwin et al., 2015; Stam & Reijneveld, 2007). Graph edges are based either on structural connectivity, defined as the strength of the axon-dendrite connections of neurons or neuronal populations or on functional connectivity, in which the interplay between the time varying signals of the nodes determines graph edges.

In this work, the focus is on functional connectivity between brain areas derived from fMRI (Sec.1.4.1). In order to obtain the nodes of the graphs, brain voxels (3D volumetric pixels of 2-3 $mm^3$) are clustered. Subsequently, the average BOLD signal of the resulting brain regions is used to define the edges of the graph. The resulting graphs undergo further graph analysis including the methods discussed above. In the end, the results are used to investigate in how far the resulting graph property distributions allow for group differentiation between healthy, MCI and AD individuals (see Sec.1.1).

However, attempting to interpret differences in the properties of fMRI-graphs between groups in a biologically meaningful way is in general difficult. On the one hand, it is difficult to discriminate between neuronal-activity based effects and vascular alterations (see Sec.1.4.1) and on the other hand, graph properties, which are based on summing up inverse correlations (as it is for example the case for shortest path metrics), have no distinct biological correlate. In addition, different ways of graph construction can lead to contradicting results, as I show in Ch.3.

### 1.4.3 Network simulation and performance measures

In order to achieve a deeper understanding of how the brain works, it would be desirable to manipulate some components in the system (e.g. the connectivity of the neurons or brain areas ) and see how its functionality will change. Due to ethical considerations and technical and/or monetary limitations, experiments based on animals or humans are often not suited to address such questions. A loophole is offered by emulating the most crucial components of the system using numerical simulations, which can then be modified in the desired way. Although such simulations allow for a systematic investigation of single components, the degree of detail the simulation offers also sets a limit to the interpretability and generalization of the results. Thus, the most fundamental question, which often can not be answered entirely is: What are the most crucial features of the system with respect to the considered scientific question? And are all necessary features included in the simulation setting in order to explain it?

Now, why not including as many features as possible in a simulation (which is not possible)? A profound reason against a detailed model is given, if the model should not only serve as an emulation tool but should also be amenable to analysis. Analytical approaches, for example, often permit predicting of the long-term behavior of the system without time-consuming simulations. In addition, after framing some system assumptions, e.g. stationarity, it is often possible to predict global statistics of the system e.g. mean or standard deviation of the firing rate of a neuronal population. Moreover, simplified model foster intuitive understanding.

Throughout this thesis, I simulate the behavior of single neurons gathered in a small network, manipulate a parameter in this model and measure the model's performance with respect to sensitivity towards a small perturbation. Finally, I provide an analytical model that substantiates the observations made and provides further insight into potential underlying mechanisms.

The reader of this thesis might wonder, how exactly neuronal behavior can be emulated and how the performance of certain tasks can be measured. The next paragraphs gives an overview on how the interaction between neurons can be simulated (partly explaining underlying mathematical equations) and what kind of models are suited for which tasks, as well as a rough sketch on how some performance measures are implemented.

The mathematical description of a neuronal network, without considering its external input or its readout (a group of neurons that is connected to the considered network), has three

primary features that have to be characterized: the connectivity of the neurons among each other, the synaptic transmission and how the input to a neuron is transferred to the output of the neuron (transfer function). For simplicity, I focus on describing neurons as a point in space (point-neuron model), neglecting the influence of morphology on its dynamics, e.g. the placing of the synapses on the dendritic tree.

The connectivity of the network describes the strengths (or other properties) of existing synapses. It can be summarized in a connectivity matrix, in which for each possible combination of a pre-post-neuron coupling a value is given. If a presynaptic neuron does not connect to a postsynatptic neuron, the entry is zero, otherwise it takes the value of the corresponding synaptic weight. Such connectivity matrices also specify a graph structure, with neurons representing the nodes of the graphs (see Sec.1.4.2). Different network connectivity structures and classes have been related to different network performances and properties. e.g. Hopfield networks (Little, 1974; Hopfield, 1982) feature associative memory and feed forward neural networks are suited for supervised learning. A short, but comprehensive overview about the different network types and their characteristics is given by Fjodor van Veen [8] .

The intrinsic dynamics of a neuron, that transfer a given input $x(t)$ at time $t$ to an output $y(t)$ can either describe an instantaneous response to an input (artificial networks), given as

$$y(t) = F[x(t)] \tag{1.1}$$

or characterize the response to the input taking into account the entire history of inputs starting from time $t_0$. Such neuron dynamics that entail 'memory' (biological networks) can be formulated in the most generic way by:

$$y(t) = F[x(t)|\forall t' \leq t], \quad y(t_0 = 0) = y_0 \tag{1.2}$$

with $y_0$ being the neuron state at time $t_0$ and $t'$ the time before $t$.

Classifying neurons based on their transfer function can yield three main groups: binary neurons (with outputs of ones or zeros/minus ones); firing rate neurons (where $y(t)$ can be seen as either the pooled firing rate of the network or the average firing rate of a single neuron across different trials); or spiking neurons. Binary neurons usually have instantaneous input responses and no 'memory'. Rate and spiking neuron models can be defined in both ways: 'memoryless' or with 'memory' (depending on the exact realization of the transfer function).

Depending on the exact type of network and neuron, different tasks, such as classification, pattern completion, prediction, sequence learning, memory formation etc. can be solved more or less successfully. So far, rate neuron models have been found to be particularly powerful for a whole range of task, e.g. classification and learning (of specific tasks), pattern recognition, data generation, denoising and compression, dimension reduction, approximations, prediction (Bishop, 2006; Schmidhuber, 2015; van den Oord et al., 2016; Hinton et al., 2006; Hornik, 1991; Connor et al., 1994). But why does the brain communicate via spikes?

[8] http://www.asimovinstitute.org/neural-network-zoo

Is it just because of energy and space saving even if the coding is less effective? Or is the underlying coding just poorly understood? New approaches extending the training results of rate neuron models to an analogues spiking network also achieve high computational power in tasks like delayed reaction, self-sustained pattern generation, long-term memory etc. (Thalmeier et al., 2016; Abbott et al., 2016). As a consequence, I assume that computation with rate neuron models is still in its fledgling stages and the advantages of communication via spikes will arise with ongoing research.

When it comes to synapses, dynamics can be describe as changes in conductance or current. Similar to the neuron's intrinsic input integration, the current synaptic input processing might include the information of previous incoming spikes (with memory). In addition, synapses can be static or plastic. Plastic synapses increase (potentiate) or decrease (depressed) their weights in accordance to the specific coupling of pre- and postsynaptic neuronal spiking and thus provide a basis for memory storage(Takeuchi et al., 2014).

After having defined the main components of a neuronal circuit, I would like to illustrate how computation can be carried out by such a network based on two basic example tasks: classification and time- series prediction.

In a classification task, different realizations of inputs are assigned to different groups according to their features. The simplest classification can be carried out by a single Perceptron, which is a binary neuron that weights the input (an array of real numbers) through its synapses and computes a step function on this sum of the weighted inputs. Thus, if the sum of the weighted input is larger then a defined threshold, it outputs a one, and a zero otherwise. Consequently, a single Perceptron allows the classification into of groups. If the input connects to more, e.g $n$ binary neurons and the groups can be separated linearly, a maximum of $n^2$ groups can be classified. If the classification task is non-linear, more layers of binary neurons have to be connected in a feed-forward manner in order to achieve an accurate classification (Gardner & Dorling, 1998). Although binary neurons are more amenable to analysis, they are often not so powerful, if the classification task is more complex. In these cases, spiking and particularly rate neuron networks are preferable.

In order to describe the second task, I will give a little background information: The successful surviving of an individual highly depends on its ability to predict the future states of changing processes in its environment based on previous observations and of course, on its ability to adapt its own behavior in the most favorable way. The prediction of future states applies to short and long time scales and different degrees of complexity. A simple example on a short time scale and low complexity might be the calculation of the movement of a fast approaching predator, such that an appropriate escape or defense behavior can be initiated. How the brain can manage such complex time-series predictions is still an open question. Specific neuronal networks with recurrent connections have been shown to be able to accomplish such tasks. An example is given by the echo state networks introduced by Jaeger 2002. Such a network consists of a big pool of recurrently connected neurons (so-called reservoir) with random but fixed connection weights. This network is recurrently connected to an output neuron. As described in Jaeger & Haas (2004), the weights of the connections from the output neuron to the reservoir are fixed and do not change during the training period. However, the connection weights from the reservoir to the output neuron are adjusted during the training

phase such, that the activity of the output neurons converges to a predefined time series. In the test phase, only the first points of the time trace are provided and the activity of the reservoir drives the activity of the output neuron such, that it predicts the time series.

These two examples have been chosen, because they demonstrate in an easily understandable manner how real-world tasks can be translated into an input-output function and, at the same time, do not require complicated network architectures. As already indicated earlier, tasks and networks can be arbitrary complex. However, even if a certain network structure provides a good substrate for solving a specific task, it does not automatically imply that the brain uses similar structures and strategies for a similar purpose. In addition, there are some tasks, that the brain manages but which cannot currently be achieved (in a biologically plausible way) with simulated network models. For example, one-shot learning (learning after one single trial, without repetition) cannot be achieved by common biological neuronal network model. It require features that are not found in the brain so far, such as on-off-switching of synapses.

In terms of Alzheimer disease (Sec.1.1), questions which, for example, can be addressed with simulation are: Which are the exact morphological alterations that cause a particular cognitive impairment e.g. memory decline and what are the underlying mechanisms? What are the mechanisms that counteract the decay in cognitive performance and how do they work? Is AD mainly a disease that occurs because homeostatic mechanisms outreach their capacity or because normal perturbations cause increased damage due to malfunctioning of homeostatic mechanisms?

## 1.5 Dealing with Alzheimer's disease and its effects on homeostatic regulation

***An introduction to the other chapters of this thesis***

In this section, I will describe how a rich homeostatic regulation interferes with disease understanding and diagnosis. Then, referring to AD, I will shortly summarize how these problems can be addressed in general, and introduce my approaches, which are explained in more detail in the following chapters.

In subsection Sec.1.2, I have given an introduction to homeostasis, explaining how the coordinated interplay of controlled variables ensures that the regulated variable stays in its physiological regime. But, what happens to the controlled variables, if the regulated variable gets pushed out of its physiological range? This pathological state can occur for several reasons, e.g a wrong set point as target or problems with controllers.

The first observation is that, considering a single controlled variable, the comparison of the marginal distributions across the subjects of two different health conditions ('healthy' and 'having a particular disease') reveals differences. But these differences might be small, such that the two distributions still highly overlap, and a full separation of the health states would

be only feasible, if the values of all involved controlled variables were projected to a high dimensional space (see Fig.1.5). I suspect, that the more controlled variables are involved in a particular homeostasis, the more likely it is, that a marginal distribution of a single controlled variable across subjects might not change a lot under disease conditions , because other controlled variables already fulfill its function. The resulting small differences in the marginal distributions across health conditions are hardly detectable.

As a result, a diagnosis of a disease, that interacts with a rich homeostatic system, is difficult, if the regulated variable is not known or measurable (if it was known and measurable, disease diagnosis could be based on the regulated variable), because all controlled variables have to be extracted and analyzed in order to clearly distinguish between health conditions. This implies that all controlled variables are measurable and that the differences in their marginal distributions are detectable. In addition, a method that allows for an analysis of the data in a higher dimensional space, has to be established.

For disease understanding, the consequences of a rich homeostasis are even more challenging. In this case, it is important to figure out, whether a change in a variable is directly caused by the pathogenic factor or whether it is part of the homeostatic response of the system. This requires knowledge of the exact interaction of the involved variables, and in particular of alternative signaling pathways, which are only activated as a consequence of the disease and not detectable under physiological circumstances. It is also possible that a disease corrupts multiple homeostatic systems on many scales, which even worsens the situations. This seems to be the case in AD (as explained in detail in Sec.1.3).

A first common step towards diagnosis and understanding a disease is to figure out, which variables change their distributions in the course of the disease. This is often done based on significance tests, in which the variables of the healthy and the diseased population are tested for having the same mean values or the same distributions. Since the nullhypothesis of such a test is that the means/distributions are equal, a small significance level indicates a low probability of this hypothesis and in turn, a high probability that the means/distributions are different. But even if the means of the two distributions are significantly different, it is important to keep in mind, that the specific experimental and data analytical settings matter and results should not be generalized, if not tested with equivalent but different analytical tools. In particular, it should be avoided to tune analysis steps aiming at the highest possible significance level, especially if the data set is small. In general, the more complex the data analysis is, the more important it is to show the robustness of the results with respect to methodical variants.

Also problematic is the conclusion that significance automatically indicates a good diagnostic power. This might be a necessary (if the data set is large) but not a sufficient criterion because the underlying distributions can still highly overlap such that for a huge span of values of the particular marker (which is a variable that indicates a disease) a high probability for both health conditions is given (see Fig.1.5, note that the histograms in this figure roughly approximate corresponding probability distributions).

So, what would be a suitable assessment criterion for a disease marker, or better, for a set of markers with overlapping distributions? Commonly, disease markers are often evaluated

based on their sensitivity (the proportion of true positives) and specificity (the proportion true negatives) of a test data set. This implies that a marker or a set of markers allow for a 'yes' or 'no' answer. This can be either achieved by assigning such a label to the health condition with highest probability or by applying machine learning algorithms that allow for a classification of the data.

Based on fMRI graphs (see Sec.1.4.1) machine learning algorithms have shown very good results with respect to psychiatric disease classification, e.g. in AD a classification power above 90% has been reported (Khazaee et al., 2015, 2017) and in Schizophrenia similar values have been reached (Çetin et al., 2016; Venkataraman et al., 2012). So why not using classifiers, if they are so successful? The disadvantage of classifiers, apart from problems involving over-fitting or a too small training data sets, is the result itself, a classification, which is not very suitable for diagnosis. The problem here is, that a diagnosis test should be combinable with other tests, which is difficult with a 'yes, she has AD' or 'no, he has not AD' decision.

In another approach a probability for the health condition of a patient given a test result is derived. This is already much more useful, because it preserves the information about the level of (un)certainty, and can be combined with other probabilities from other tests in order to calculate the final probability. In this context, a suitable set of disease markers should result in probability distributions that will in almost all cases correctly assign a high probability to one of the possible health conditions and a low probability to the others.

In the next chapters of this work, I focus on two systems on different scales, in which AD is likely to have an impact on their homeostatic regulation: the first system is a metascale network of functionally connected brain areas (see Ch.2 and Ch.3) and the second is a mesoscale network of spiking neurons (see Ch.4). At the mesoscopic scale of individual neurons, the firing rate is most likely a regulated variable in AD (see Sec.1.3). With respect to the meta-scale network, it is not known yet, in how far functional whole-brain connectivity underlies direct homeostatic control. First hints for this type of global homeostasis has been given in the context of simulating the generation of seizures in an EEG model (Chakravarthy et al., 2009). Even if functional connectivity is not subject to direct homeostatic control, the activities of individual areas underlie homeostatic control, and homeostatic misregulation in these areas will affect functional connectivity.

In Ch.2 I will describe, how variables with highly overlapping distributions across health conditions can be used for disease diagnosis and how, for that purpose, the statistical model is constructed. In this case the variables are derived from functional connectivity measures. For some psychiatric diseases such as Schizophrenia or AD, in which large brain structures are affected, the functional connectivity is likely to be altered in any sort. It is a welcome potential disease marker due to the non-invasive nature of MRI scans and and the already well established usage of the fMRI scanners in the western medical infrastructure.

The diagnostic power of a set of markers highly depends on the suitability of the underlying statistical model, which is based on assumptions made on the data. Unfortunately, these assumptions are often not really known and the statistical model is chosen in a rather ad-hoc manner. This profound problem will be particularly addressed in Ch.2 of this work, which describes how probabilities or rather likelihoods can be derived in a step by step

manner starting from first principle. Therefore, it explains how statistical models can be constructed based on only a few but simple assumptions such as partial exchangeability and sufficient statistics and demonstrates how the exact realization of these assumptions shapes the final model selection. In addition, it explains how the different models can be compared among each other with respect to a possible medical usage and how side effects of a possible treatment can be accounted for by means of a utility function.

The investigation of the different statistical models, their underlying assumption and practicability in Ch.2 forms the basis of the statistical analysis with which fMRI data is analyzed in Ch.3. In Ch.3 I apply different graph construction methods (see Sec.1.4.2) to functional MRI data (see Sec.1.4.1) and compare the resulting graph properties with respect to their diagnostic power in AD.

As described in Sec.1.1, AD is a disease that destroys the brain in many respects at different times, primarily triggering synaptic dysfunction, but also other neuronal activity independent changes such as vascular alterations. Both aspects are captured by fMRI (compare Sec.1.4.1). This suggests that alterations of the functional connectivity in all disease stages are very likely. The distributions of the graph properties are, however, overlapping, as I show in Ch.3. This overlap may be very prominent at the beginning of disease manifestation, because the disease has not driven the system far from its healthy state. However, as explained in Sec.1.1, an early diagnosis in AD is essential. In order to address this issue, I study the graph properties not only of AD patients, but also of individuals with mild cognitive impairment (MCI), a potential pre-stage of AD.

In this context, I also analyze the role of significant differences in the distributions of the graph properties such as the clustering coefficient. In particular I demonstrate that different graph construction methods can lead to contradicting significant relationships of graph properties across health conditions.

FMRI BOLD signals contain both a vascular and a neuronal activity component (see Sec.1.4.1). A more detailed picture, allowing for discriminating the underlying processes can be obtained by calibrated fMRI. So far, only a single study of calibrated fMRI has been published in AD research (Lajoie et al., 2017). This study reports a hypomethabolism in parietotemporal regions reflecting decreased neuronal activity. Thus, in early states of AD, changes in the BOLD signal seem to reflect altered neuronal activity. If altered graph properties mainly reflect altered neuronal activity and not vascular alterations, it would be desirable to figure out whether the cause of these alterations is of homeostatic nature or a direct consequence of disease pathology. A first step towards addressing this question, is the investigation of disease-caused modifications and possible compensation mechanisms in local networks. The reason is that changes in the activity of local networks will most likely alter the functional activity of the global network.

In Ch.4 of this work, I investigate the effect of AD and firing-rate homeostatis on the dynamical and computational properties of recurrent neuronal networks. I conjecture, that if homeostasis can recover the computational performance and the activity statistics of a small network of spiking neurons, it will also recover the functionality of a large-scale network. For reasons of simplicity, I restrict myself to implement AD as a loss of excitatory-excitatory

synapses and homeostasis as an upscaling the remaining excitatory-excitatory synapses. I primarily focus on the network's sensitivity to a small perturbation.

In the last chapter (Ch.5), I summarize the findings of the studies presented in Chs.2-4 and discuss, how they integrate with the results of previous work and give a short outlook.

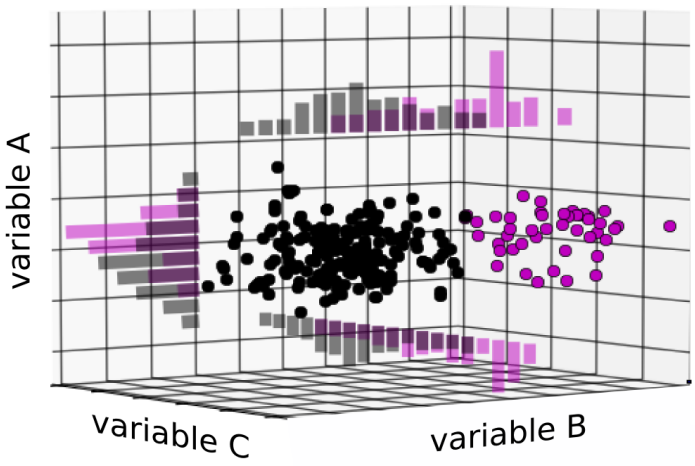


**Figure 1.5: Sketch of a homeostatic system comprising three controlled variables.** A regulated variable, that is controlled by only three controlled variables (variable A, B and C), changes its value in the course of a disease (healthy: black, diseased: purple). Each dot represents an individual. Even if the marginal distributions highly overlap (black and purple histograms), a separation of the two different health conditions based on the three controlled variables is possible, if the values are embedded into a three dimensional space (joint distribution). In a living organism we usually have many more controlled variables, but only some of them can be observed. The combined information of a subset of controlled variables, represented by the joint distribution, can improve the separation of the health conditions.

# Chapter 2

# Inferring health conditions from fMRI-graphs

Automated classification methods for disease diagnosis are currently in the limelight, especially for imaging data. Classification does not fully meet a clinician's needs, however: in order to combine the results of multiple tests and decide on a course of treatment, a clinician needs the likelihood of a given health condition rather than binary classification yielded by such methods. We illustrate how likelihoods can be derived step by step from first principles and approximations, and how they can be assessed and selected, using fMRI data from a publicly available data set containing schizophrenic and healthy control subjects, as a working example. We start from the basic assumption of partial exchangeability, and then the notion of sufficient statistics and the "method of translation" (Edgeworth, 1898) combined with conjugate priors. This method can be used to construct a likelihood that can be used to compare different data-reduction algorithms. Despite the simplifications and possibly unrealistic assumptions used to illustrate the method, we obtain classification results comparable to previous, more realistic studies about schizophrenia, whilst yielding likelihoods that can naturally be combined with the results of other diagnostic tests.

## 2.1 Introduction

A 29-year-old man seeks medical advice because he finds himself in a very confused state. The clinician, after listening to the complaints of the patient, identifies some diseases that would account for the symptoms. However, the presentation is not clear cut, and treatment for some of the potential conditions have significant side effects. To come to a decision on the best course of action, the clinician decides to perform the differential diagnosis in a mathematically sound manner (Sox et al., 2013), first assigning an initial probability for the patient's being healthy or having each of the potential conditions, taking into account age, sex, familial factors, symptoms, a psychological evaluation, the incidence of the disease, and

similar prior information:

$$\mathrm{P}(\text{health condition} \mid \text{prior info}). \tag{2.1}$$

Then she orders one or more diagnostic tests to make a better informed assessment of the probabilities of the considered diseases. Among these tests she orders a structural and functional magnetic resonance imaging (MRI) scan. The advantage of MRI lies in the non-invasive monitoring of brain structure and activity; the structural image (sMRI) is used to exclude morphological changes in the brain such as tumours, and the functional imaging (fMRI) can provide information about changes in brain activity.

With the results of the tests and of the sMRI and fMRI, the clinician updates her initial or prior probability to a "post-test" or posterior probability based on the results, according to Bayes's theorem:

$$\overbrace{\mathrm{P}(\text{health condition} \mid \text{results of all tests} \wedge \text{prior info})}^{\textit{post-test probability}} \propto \overbrace{\mathrm{P}(\text{health condition} \mid \text{prior info})}^{\textit{initial probability}} \times$$

$$\textit{likelihoods} \begin{cases} \mathrm{P}(\text{result of first test} \mid \text{health condition} \wedge \text{prior info}) \times \\ \mathrm{P}(\text{result of second test} \mid \text{health condition} \wedge \text{prior info}) \times \\ \cdots \times \\ \mathrm{P}(\text{result of sMRI} \mid \text{health condition} \wedge \text{prior info}) \times \\ \mathrm{P}(\text{result of fMRI} \mid \text{health condition} \wedge \text{prior info}) \end{cases} \tag{2.2}$$

where "$\wedge$" denotes the logical conjunction ("and"), and we have reasonably assumed that the result of each test does not depend on those of the other tests, i.e. that their likelihoods are independent (Jaynes, 2003, Sec. 4.2; Sox et al., 2013, Sec. 4.7).

In the update formula above, the initial probability is assessed by the clinician. To calculate the post-test probability she needs the probabilities for each test result conditional on the health condition, either "healthy" or "presumptive disease". These probabilities are called the *likelihoods for the health condition* in view of each test. The term "likelihood" has its standard technical meaning in the present work: the probability of a proposition $A$ given $B$ is $\mathrm{P}(A \mid B)$, while the likelihood of $A$ in view of $B$ is $\mathrm{P}(B \mid A)$, i.e., $A$ appears in the conditional (Jaynes, 2003, Sec. 4.1; Good, 1950, Sec. 6.1). A proposition can have high probability but low likelihood and vice versa. Probabilities, not likelihoods, are what we base our decisions on.

The final, post-test probability is necessary to the clinician to decide upon a course of action (Sox et al., 2013, Ch. 6; Goodman, 1999; Murphy, 2012, Sec. 5.7); for example, to treat the patient according to one or another specific treatment, to dismiss him, or to order more tests. To make such a decision the clinician will combine her post-test probabilities for the health conditions with a utility table (a reminder of decision theory is given in Sec. 2.4.3).

In the following we assume that one of the presumptive diseases the clinician has in her mind is schizophrenia. Although currently, MRI does not play a role in a diagnosis of schizophrenia,

there are substantial efforts to develop such analyses for this purpose (Silva et al., 2014). In this work, we focus on the diagnosis of this particular disease simply as a concrete worked example, to demonstrate how results of a diagnostic test, in this case results from function MRI imaging, can be incorporated in the diagnostic process in a principled fashion.

In short, we address the question: **how can we assign a numerical value to the likelihood**

$$\mathrm{P}(\text{fMRI result} \mid \text{health condition} \wedge \text{prior info}). \tag{2.3}$$

**of each health condition ('healthy' or 'schizophrenic') in view of the fMRI result?** We will propose an answer that can be applied for any brain disease.

To this end it is useful to mark out some features of the approach presented so far:

I.**Modularity.** The update formula (2.2) combines evidence from different tests, and this combination does not need to be done at once. The clinician can multiply her initial probability by the likelihood from the first test, normalize, and thus obtain a "post-first-test" probability. Later she can multiply this probability by the likelihood from the second test, normalize, and thus obtain a post-second-test probability; and so on with any number of other tests, a number that the clinician does not need to fix in advance. She can therefore store the value of the likelihood from the fMRI result, to later combine it with new likelihoods from future tests to form a new, better-informed post-test probability.

II.**Decision-theoretic character.** The clinician's final goal is not simply a healthy/schizophrenic classification, but a *decision* upon a course of action about the patient (Sox et al., 2013, Chs. 6, 7; Jaynes, 2003, Chs. 13, 14; Raiffa & Schlaifer, 2000). This distinction is important: for example, a treatment without contraindications might be recommended even if there is only a 10% probability that the disease is present; or a dangerous treatment might be recommended only if there is a 90% probability that the disease is present.

The modularity of the present approach extends to the decision stage, because the post-test probability can be used with different decisions and utilities, which can also be updated later on. For example, after beginning a treatment the clinician happens to read about a new kind of treatment, having new benefits and contraindications. Using the post-test probabilities she already has, she may re-evaluate her decision using an updated utility table that includes the new treatment.

III.**Incomplete knowledge.** In general, we lack a complete biological understanding of the relation between brain activity and the health condition under study. In this case, the likelihoods can only be assessed by relying on examples of known *health condition–fMRI data* pairs, usually called a training or calibrating dataset. Moreover, this training dataset is often very small.

IV.**High dimensionality.** The fMRI data are positive-valued vectors with $10^7$–$10^8$ components or more (Lindquist, 2008). This high dimension impacts the calculation of likelihoods and probabilities.

The first two points above are great advantages of the present approach, and also the reasons

why it cannot be based on machine learning algorithms for deterministic classification; such methods give the clinician a dichotomous, "healthy/schizophrenic" answer, with no associated uncertainty. This answer cannot be used by the clinician to weigh the benefits and risks of different courses of action, an assessment which needs the probabilities of the health conditions. Probabilistic algorithms, on the other hand, are not flexible for combining evidence: they give a probability for the health condition, not a likelihood; and only the latter can be combined with the likelihoods from other tests, or stored for later reuse and combination.

We therefore approach the question of assigning the likelihoods (2.3) by means of the probability calculus, the same calculus from which Eq. (2.2) is derived. What we will do is in essence no different from current Bayesian statistical analyses and modelling; but we would like to emphasize some aspects of this modelling that are usually left in the background. The probability calculus can be regarded as the extension of formal logic (truth calculus) to plausible inference (Jeffreys, 2003; Jaynes, 2003; Hailperin, 1996), a view also supported in medicine (Greenland, 1998; Maclure, 1998; Goodman, 1999), which has been proven with increasing rigour by Koopman (1940b; 1940a; 1941), Cox (1946; 1961; 1979), Pólya (1949; 1968), and many others (Horvitz et al., 1986; Paris, 2006; Halpern, 1999; Snow, 1998; Dupré & Tipler, 2009; Terenin & Draper, 2017). The derivation of a probability proceeds much like an "axioms $\rightarrow$ logic rules $\rightarrow$ theorem" derivation in formal logic: one starts from the probabilities of some propositions, and by applying the probability rules, arrives at the probability of the desired proposition, Eq. (2.3) in our case.

We will show this procedure step by step in the case of our problem, in order to expose where assumptions and approximations enter the derivation. These may be improved by other researchers, or replaced by different ones when the method is applied to a different problem. Our discussion is inspired by Mosteller & Wallace's (Mosteller & Wallace, 1963) brilliant, thoughtful analysis of a statistically similar problem in a very different context.

The approach we follow deals naturally with the four points listed above. The small size of the training dataset, point III. above, is not an issue because the probability calculus allows for training datasets of any size. In fact, the calculus allows us to continuously update our inferences given new training data, making our inferences more and more precise and less likely to be affected by outliers.

The unmanageable size of our data space, point IV. above, will force us to make auxiliary assumptions that will translate into the choice of a reduced data space, discussed in Sec. 2.2.3, and into the use of parametric statistical models, discussed in Sec. 2.2.4. Regarding the latter, we will emphasize that assumptions about relevant and irrelevant information in our data may translate into mathematical statistical models. It is often difficult to relate biophysical considerations about quantities measured in the brain to the shape of a probability distribution, especially in multidimensional quantities. The notion of *sufficient statistics* (Dawid, 2013; Bernardo & Smith, 2000, Ch. 4; Lindley, 2008, Sec. 5.5; Diaconis & Freedman, 1981; Cifarelli & Regazzini, 1982; Lauritzen, 1988; Kallenberg, 2005), discussed in Sec. 2.2.4.2, is a helpful bridge between biophysical considerations and probability distributions. The idea is that it may be easier for us to conceive a connection between biophysical considerations and some special statistics of our measurements, than between biophysical considerations and an

abstract multidimensional distribution function. This "translation" is powerful, because if one finds the assumptions about relevance or irrelevance of some data unreasonable, one can then make different assumptions, resulting in a different statistical model.

Models inspired by sufficient statistics – especially their comparison and selection – can nevertheless be computationally demanding owing to the multidimensional integrals in their formulae, even when these are addressed by modern numerical methods such as Monte Carlo (MacKay, 2003, Ch. IV; Murphy, 2012, Chs. 23–24). In the present study we shall use analytically tractable statistical models, but availing ourselves of Edgeworth's "Method of Translation" (1898), (Johnson, 1949; Mead, 1965): the simple but potentially very fertile idea of transforming a quantity into a normally distributed one, discussed in Sec. 2.2.4.3.

The combined choices of reduction of the data space, of sufficient statistics, and of transformations into normal variables, lead to a variety of possible models and likelihoods to be used by the clinician. Which is the "best" one? We discuss several criteria for choice in Sec. 2.2.5, settling on one based on expected utility. We also briefly discuss the remarkable observation that common Bayesian criteria based on weight of evidence and Bayes factors (Jeffreys, 2003, Chs. V, VI, A; Good, 1950; MacKay, 1992a; Kass & Raftery, 1995) for the fMRI data gives results opposite to those of the expected-utility criterion.

In this article, we will calculate the likelihoods for the health conditions, Eq. (2.3), and assess the models according to the following steps. First, in Sec. 2.2.1, we briefly discuss schizophrenia and the use of fMRI to diagnose it, introducing a concrete dataset of fMRI data for schizophrenic and healthy patients. We then show that a simple and natural assumption, called *exchangeability*, would lead to a unique value of the likelihoods (2.3), if the training dataset were large enough (Sec. 2.2.2). However, with a small training dataset we must face two problems: unmanageably large dimensions of the data space, and the need to specify prior beliefs, also involving functions on infinite-dimensional spaces.

To solve the first problem, in Sec. 2.2.3 we assume that information adequate for our health inference can be found in a reduced data space of the fMRI, which we construct from time correlations between groups of voxels. To solve the second problem we introduce parametric statistical models in Sec. 2.2.4 using the notions of *sufficient statistics* and of transformation into normal variables, mentioned above. We discuss how these models learn from the data and select three models as possible candidates. We then consider several criteria to select one of the three models against our data, as an example, and discuss how a more realistic assessment could be made in a real application (Sec. 2.2.5 ). We conclude with a discussion (Sec. 2.3) on how the choice of sufficient statistics and prior probabilities could be improved, and on the relation to machine-learning methods.

Our statistical terminology and notation follow ISO standards (ISO, 2009, 2006).

## 2.2 Results

### 2.2.1 Selection of clinical use case and fMRI-data acquisition

Schizophrenia is a psychiatric disorder that comprises various symptoms that are categorized into positive (e.g. hallucinations), negative (e.g. loss of motivation) and cognitive (e.g. memory impairment) disease patterns. A common disease cause for all these widespread symptoms is still unknown. Functional magnetic resonance imaging (fMRI) has been used to gain insight into modifications in functional connectivity in this disease. In the resting state, functional connectivity is measured either by asking the subject to fulfil a certain task or at rest, instructing the subject to think about nothing specific but not fall asleep. In this condition, both increased and decreased functional connectivity have been reported in the default mode network, although the hyperactivity seems to be reported more often (Hu et al., 2017). Moreover, widespread connectivity changes in the dorsal attention network and the executive control network have been detected (Woodward et al., 2011; Yu et al., 2016).

Beyond these individual sub-networks, many studies have found profound changes in macroscopic brain structures, e.g. a shrinkage of whole brain and ventricular volume, reduced gray matter in frontal, temporal cortex and thalamus, and changes in white matter volume in frontal and temporal cortex (Shenton et al., 2010; Ellison-Wright & Bullmore, 2009, 2010). Since both gray matter loss and white matter changes are found, it is reasonable to conclude that not only the intrinsic activity of single areas is modified, but also the interplay of different brain areas, in particular in frontal and temporal cortex. It has been argued that these alterations in long range connectivity are responsible for a range of disease symptoms that are not attributable to single areas (Friston & Frith, 1995). Taking this disconnect hypothesis as a starting point, we can reach the working hypothesis that these changes are also reflected in the functional activity of the brain, and that fMRI images can be used to distinguish schizophrenic from healthy patients. We therefore conclude that schizophrenia is an appropriate condition to demonstrate our approach.

We requested data of schizophrenic and healthy patients from Schizconnect[1], a virtual database for public schizophrenia neuroimaging data. In our request we asked for resting state T2*-weighted functional (rfMRI) and T1-weighted structural magnet resonance images (MRI) from patients participating in the COBRE study either with no known disorder or diagnosed as schizophrenic according to the Diagnostic and Statistical Manual of Mental Disorders (DSM) IV, excluding schizoaffective disorders. In the COBRE study, the voluntary and informed participation of the subjects was ensured by the institutional guidelines at the University of New Mexico Human Research Protections Office. The ensuing dataset comprised 91 healthy patients and 74 schizophrenic patients. Out of these we randomly selected 54 healthy and 49 schizophrenic subjects, to permit demonstration our method on a small dataset with unequal group size. A detailed description on the exact experimental design and the MRI scanning is provided by Çetin et al. (2014).

[1] http://schizconnect.org/

### 2.2.2 Calculation of probabilities: exchangeability

Let us describe our context more precisely and set up some mathematical notation. We have:

- A number of possible health conditions, in our example healthy (H) and schizophrenic (S). The variable $c$ denotes health condition.
- A space of possible fMRI data. They are vectors with $10^7$–$10^8$ or more positive components (Lindquist, 2008). The variable $\boldsymbol{f}$ denotes an fMRI result.
- A set of $n$ patients, labelled in some way, the variable $i$ denoting their labels. These labels may reflect information about the times the patients were examined, or about their geographical location. This possibility is important in the considerations to follow. In our study $n = 104$.
- Knowledge of the health condition and of the fMRI result of each patient. Let us use the propositions
$$\begin{aligned} C_i^c &\coloneqq \text{“Patient } i \text{ has health condition } c\text{”}, \\ F_i^{\boldsymbol{f}} &\coloneqq \text{“The fMRI of patient } i \text{ gives } \boldsymbol{f}\text{”}, \end{aligned} \tag{2.4}$$
the latter to be understood within a very small interval $(\boldsymbol{f}, \boldsymbol{f} + \mathrm{d}\boldsymbol{f})$. In our study we have $n_{\mathrm{H}} = 55$ healthy and $n_{\mathrm{S}} = 49$ schizophrenic patients.
For brevity we denote by $C^c$ the conjunction of the propositions $C_i^c$ for all patients having health condition $c$, i.e. our knowledge about which patients have that health condition; and analogously for $F^c$. By $D^c$ we denote all data about patients with health condition $c$; by $D$ we denote all our data.
- An imaginary patient, labelled “0”, whose fMRI result $\boldsymbol{f}$ is known, but whose health condition is not.
- Other pre-test information, denoted by $I$; for example the clinician’s initial diagnosis of the health condition of patient 0, and the results of any other diagnostic tests.
- The probabilities $(\hat{p}_{\mathrm{H}}, \hat{p}_{\mathrm{S}})$ for the health condition of patient 0, conditional on the pre-test information, including the results from other tests. We call these *pre-test probabilities*. Note that they may differ from the *initial* probabilities of Eqs. (2.1)–(2.2), because they may include the likelihoods from other tests.
- A set of decisions about the patient 0 and their utilities conditional on the patient’s health condition. We shall simply consider two decisions: dismiss (D) or treat (T). See Sec. 2.4.3 for a summary of decision theory.

Our goal is to assign numerical values to the likelihoods for the health conditions (2.3): the conditional probability distribution that the fMRI result of patient “0” is $\boldsymbol{f}$ given the health condition of that patient and all other data. In our notation,

$$\begin{aligned} & \mathrm{p}(F_0^{\boldsymbol{f}} \mid C_0^c \wedge C_1^{c_1} \wedge F_1^{\boldsymbol{f}_1} \wedge \cdots \wedge C_n^{c_n} \wedge F_n^{\boldsymbol{f}_n} \wedge I)\, \mathrm{d}\boldsymbol{f} \\ \text{or just} \quad & \mathrm{p}(F_0^{\boldsymbol{f}} \mid C_0^c \wedge D \wedge I)\, \mathrm{d}\boldsymbol{f}. \end{aligned} \tag{2.5}$$

A natural assumption helps us to restrict the values the distribution above may have. Within a group of patients *having the same health condition* we assume that the probability that a patient shows a particular fMRI $\boldsymbol{f}_i$ does not depend on the particular value of the patient’s

label $i$, no matter how many patients we have or may later add in that health group. This assumption is called *partial exchangeability* (de Finetti, 1938; Diaconis & Freedman, 1981; Aldous, 1985; Diaconis, 1988). If the labels carry e.g. temporal or geographical information, partial exchangeability means that we do not expect to observe particular kinds of fMRI results more often in the future than in the past, or more frequently in one location than in another. As a concrete example: fix three possible fMRI results $\boldsymbol{f}_1$, $\boldsymbol{f}_2$, $\boldsymbol{f}_3$ (each is a vector with $10^7$–$10^8$ positive components) and consider the fMRI tests of three schizophrenic patients: say, one from five years ago in Germany, one from last week in Scotland, and one to be done six months from now in Italy. Partial exchangeability means that the probability that the German patient's test gave $\boldsymbol{f}_1$, the Scottish's gave $\boldsymbol{f}_2$, and the Italian's will give $\boldsymbol{f}_3$, is numerically equal to the probability that the German's gave $\boldsymbol{f}_2$, the Scottish's $\boldsymbol{f}_3$, and the Italian's will give $\boldsymbol{f}_1$; and likewise for all six possible permutations of the three results. Keeping the same fixed fMRI results $\boldsymbol{f}_1$, $\boldsymbol{f}_2$, $\boldsymbol{f}_3$, we now consider three healthy patients instead, who may also live in different times and places. Partial exchangeability means that also in this case the values of the six possible joint probabilities obtained by permutation must all be equal – but this value can be different from the one for the schizophrenic patients considered before. Hence the term "partial": we can freely exchange the joint results within the schizophrenic group and within the healthy group without altering their probabilities, but not across groups. This assumption extends in an analogous way to more patients.

The assumption of partial exchangeability might not be completely true when we consider geographical or epochal differences, but we may still consider it as a good approximation. We are not making any exchangeability assumptions about the probabilities of the health conditions of our patients, though, because the incidence of a disease does often change with time and can depend heavily on geographical location.

To express partial exchangeability mathematically, suppose that the patients $i = 1, 2, 3, \dots$ have health condition $c = \mathrm{H}$ and the patients $i' = 1', 2', 3', \dots$ health condition $c = \mathrm{S}$. Then the joint distribution for their fMRI results satisfies

$$\mathrm{p}\Big(\bigwedge_i F_i^{\boldsymbol{f}_i} \bigwedge_{i'} F_{i'}^{\boldsymbol{f}_{i'}} \Big|\, \bigwedge_i C_i^{\mathrm{H}} \bigwedge_{i'} C_{i'}^{\mathrm{S}} \wedge I\Big) = \mathrm{p}\Big(\bigwedge_i F_i^{\boldsymbol{f}_{\pi(i)}} \bigwedge_{i'} F_{i'}^{\boldsymbol{f}_{\pi'(i')}} \Big|\, \bigwedge_i C_i^{\mathrm{H}} \bigwedge_{i'} C_{i'}^{\mathrm{S}} \wedge I\Big)$$
$$\text{for all permutations } \pi \text{ of } \{i\} \equiv \{1, 2, \dots\}, \text{ and all permutations } \pi' \text{ of } \{i'\} \equiv \{1', 2', \dots\}. \tag{2.6}$$

The assumption of partial exchangeability is simple and quite natural – and very powerful: it implies, as shown by de Finetti ((1938); Diaconis, 1988, Sec. 3; Bernardo & Smith, 2000, Sec. 4.6), that the joint distributions above must have the form

$$\mathrm{p}\Big(\bigwedge_i F_i^{\boldsymbol{f}_i} \bigwedge_{i'} F_{i'}^{\boldsymbol{f}_{i'}} \Big|\, C \wedge I\Big) = \iint \Big[\prod_i q_{\mathrm{H}}(\boldsymbol{f}_i)\Big] \Big[\prod_{i'} q_{\mathrm{S}}(\boldsymbol{f}_{i'})\Big]\, \mathrm{p}(q_{\mathrm{H}}, q_{\mathrm{S}} | \, I)\, \mathrm{d}q_{\mathrm{H}}\, \mathrm{d}q_{\mathrm{S}} \tag{2.7}$$

where $q_{\mathrm{H}}$, $q_{\mathrm{S}}$ are distributions over the possible values of $\boldsymbol{f}$, and $\mathrm{p}(q_{\mathrm{H}}, q_{\mathrm{S}} | \, I)$ is a "hyperdistribution" over such distributions, determined by the assumptions $I$. The double integral (which can be understood as a generalized Riemann integral: Lamoreaux & Armstrong,

1998; Swartz, 2001; Kurtz & Swartz, 2004), is over all distributions $q_{\mathrm{H}}$, $q_{\mathrm{S}}$. In other words, de Finetti's theorem say that the joint probability distribution for the fMRIs of healthy and schizophrenic patients can be seen as the product of independent distributions, identical for healthy cases and identical for schizophrenic cases but different for the two cases, mixed over all possible such pairs of distributions with weight $\mathrm{p}(q_{\mathrm{H}}, q_{\mathrm{S}}|\, I)$.

As a very cursory example, suppose we want the joint probability that a healthy patient has fMRI result $\boldsymbol{f}$ and a schizophrenic one $\boldsymbol{f}'$. De Finetti's formula tells us to consider all possible distributions over positive vectors. As usual with infinities, "all" must be made precise by specifying a topology (for details see e.g. de Finetti, 1938; Diaconis & Freedman, 1981; Aldous, 1985; Diaconis, 1988); but intuitively these distributions comprise, e.g., multivariate truncated normals, gammas, exponentials, truncated Cauchys… and innumerable distributions that we can imagine and don't have a specific name for; all with their possible parameter values. De Finetti's formula tells us to choose one distribution $q_{\mathrm{H}}$, from all those possible ones, for the healthy case and one $q_{\mathrm{S}}$ for the schizophrenic case, and to attach a weight to this pair, $\mathrm{p}(q_{\mathrm{H}}, q_{\mathrm{S}}|\, I)$; then to calculate this pair at the values $\boldsymbol{f}$, $\boldsymbol{f}'$ and multiply them: $q_{\mathrm{H}}(\boldsymbol{f}) \times q_{\mathrm{S}}(\boldsymbol{f}')$. Then we consider a new pair of distributions, attach a weight to them, and again multiply their values at $\boldsymbol{f}$ and $\boldsymbol{f}'$. And so on, until all possible pairs are considered. Finally we calculate the sum of all such products, weighted accordingly: $\int q_{\mathrm{H}}(\boldsymbol{f}) q_{\mathrm{S}}(\boldsymbol{f}')\, \mathrm{p}(q_{\mathrm{H}}, q_{\mathrm{S}}|\, I)\, \mathrm{d}\boldsymbol{f}\, \mathrm{d}\boldsymbol{f}'$.

The generalization of the formulae above to more than two health conditions, or when only one health condition is considered, is straightforward.

As the cursory example above made quite clear, an integral over probability distributions is a mathematically complicated object (cf. Ferguson, 1974) and may not seem a great advancement in assigning a value to the distribution (2.5). In defence of de Finetti's formula we must say that it is completely manageable with discrete data spaces and provides a great insight in the way we reason about probability in relation to repeated events (de Finetti, 1937; Lindley & Phillips, 1976; Kingman, 1978; Koch & Spizzichino, 1982; Dawid, 2013; Bernardo & Smith, 2000, Sec. 4.2). It also has several important consequences for our inference, which we now discuss.

Using de Finetti's formula and the definition of conditional probability we can rewrite our goal plausibility (2.5) as

$$\mathrm{p}(F_0^{\boldsymbol{f}} |\, C_0^c \wedge D \wedge I) = \int q_c(\boldsymbol{f})\, \mathrm{p}(q_{\mathrm{H}}, q_{\mathrm{S}}|\, D \wedge I)\, \mathrm{d}q_{\mathrm{H}}\, \mathrm{d}q_{\mathrm{S}} \tag{2.8a}$$

$$\text{with} \quad \mathrm{p}(q_{\mathrm{H}}, q_{\mathrm{S}}|\, D \wedge I) = \frac{\left[\prod_i q_{c_i}(\boldsymbol{f}_i)\right] \mathrm{p}(q_{\mathrm{H}}, q_{\mathrm{S}}|\, I)}{\int \left[\prod_i q_{c_i}(\boldsymbol{f}_i)\right] \mathrm{p}(q_{\mathrm{H}}, q_{\mathrm{S}}|\, I)\, \mathrm{d}q_{\mathrm{H}}\, \mathrm{d}q_{\mathrm{S}}}. \tag{2.8b}$$

The latter is called posterior distribution since it is conditional on all data $D$.

Excluding pathological prior distributions (Diaconis & Freedman, 1986), this posterior distribution becomes more and more concentrated on two particular distributions $\left(\hat{q}_{\mathrm{H}}, \hat{q}_{\mathrm{S}}\right)$, fully determined by the data, as our data $D$ comprise a larger and larger number of patients. This concentration occurs independently of the original shape of the distribution $\mathrm{p}(q_{\mathrm{H}}, q_{\mathrm{S}}|\, I)$. In

this limit our probability distribution (2.5) becomes

$$\mathrm{p}(F_0^{\boldsymbol{f}} \mid C_0^c \wedge D \wedge I) \approx \hat{q}_c(\boldsymbol{f}) \tag{2.9}$$

with $\hat{q}_c$ completely determined by the data $D$. This would solve our plausibility assessment (2.5).

De Finetti's formula therefore tells us also the theoretical limit by which the pre-test probability for the health condition of the patient, $\mathrm{P}(C_0^c \mid D \wedge I)$, can be improved by the fMRI result. For example, if $\hat{q}_c(\boldsymbol{f})$ is more or less uniform in $\boldsymbol{f}$ or has the same peaks in $\boldsymbol{f}$ for each $c$, then the fMRI is of no use for discriminating the health condition of the patient. This result follows mathematically from the assumption of exchangeability (2.6) and the rules of probability calculus, hence no amount of ingenuity could overcome this limit.

In our case the amount of data $D$ is not enough to allow the use of the approximation (2.9). We should use the general formula (2.8), but it is unwieldy in two respects. First, the fMRI result $\boldsymbol{f}$ of a patient is a positive-valued vector with $10^7$–$10^8$ components or more (Lindquist, 2008), so the distributions $q_{\mathrm{H}}$, $q_{\mathrm{S}}$ are highly multidimensional. Second, the formula asks us to consider in principle *all* such distributions, as explained in the example above.

We tame this double unwieldiness in two ways. First, it is conceivable that not all information contained in the fMRI result $\boldsymbol{f}$ of a patient be relevant to discriminate the patient's health condition $c$. The integral in Eq. (2.8), if it could be performed, would automatically winnow out the relevant information (Jaynes, 2003, Ch. 17), possibly reducing the problem to a lower-dimensional set in the space of fMRI data. Being unable to perform the integral, we must try to apply heuristics based on our understanding of the target conditions and perform such dimensional reduction by hand. For example, by employing the hypothesis that in schizophrenic patients the time correlation between brain regions is altered with respect to healthy ones. We can thus address the first problem by reducing fMRI data $\boldsymbol{f}$ to a manageable set of graph properties $\boldsymbol{f}$, and applying our inference directly on these, as explained in the next section.

Second, we entertain a working assumption about which features of our graph data $\{\boldsymbol{f}_i\}$ from a set of patients are relevant for inferences about new patients. We can, for example, assume that only the first and second moments are relevant for making predictions about the graph quantities to be observed in the new patients; these moments are then called *sufficient statistics*. Assumptions of this kind reduce the infinite-dimensional space of all possible distributions $(q_{\mathrm{H}}, q_{\mathrm{S}})$ to a finite-dimensional space of a parametric family of distributions of exponential type, as explained in Sec. 2.2.4.

### 2.2.3 Trimming the data space: functional connectivity

The preprocessed functional image in standard space, in which the activity for each voxel is recorded, consists of approximately $10^{17}$ time series of $140$ time points. We reduce this huge data space by the following steps.

First, we consider only the activity of the voxels belonging to the $94$ regions defined by the lateral cortical Oxford atlas (see Sec. 2.2.1 and Desikan et al., 2006) and average the activity of all voxels in a region (details in Sec. 2.4.1). Second, we measure the functional connectivity defined by the Pearson correlation coefficient between pairs of regions, obtaining $94 \times 93/2 = 4371$ connectivity weights. This is still a considerable data space for our computational resources, so we select $d = 40$ connectivity weights that exhibit the greatest difference in their connection weight average across the schizophrenic and healthy groups.

The resulting distributions for four of these connectivity weights are depicted in Fig. 2.1. Note that for each considered brain connection, the histograms of the healthy group and the schizophrenic group display significant overlap, such that none of them could be used in isolation to reliably discriminate between two groups.

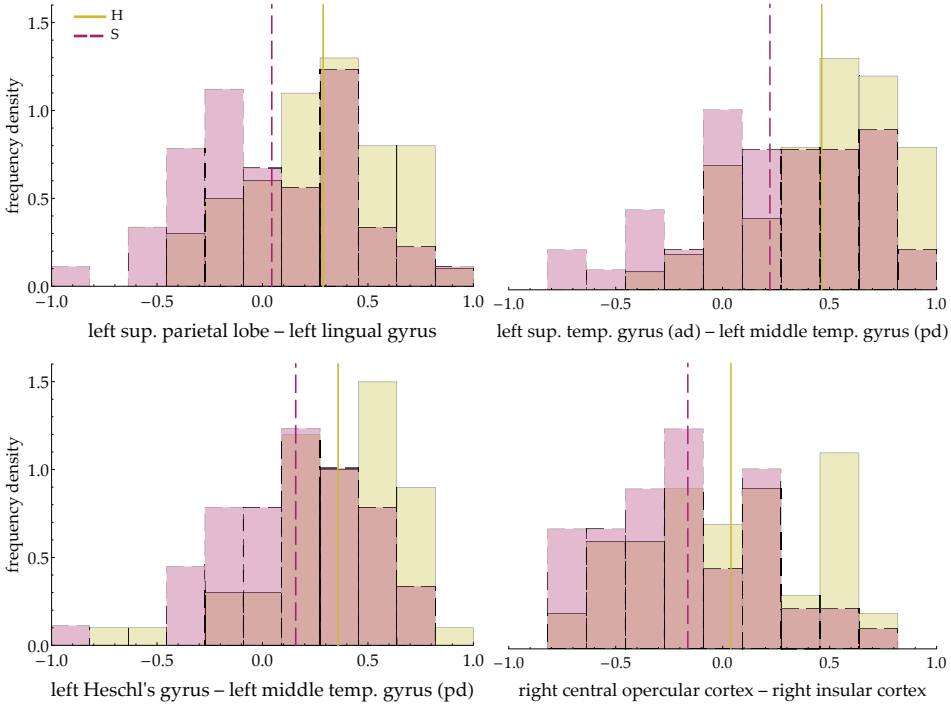


**Figure 2.1: Distributions of functional connectivities of schizophrenic and healthy patients.** Normalized density histograms of connectivity weights for healthy (H, yellow, solid) and schizophrenic (S, red, dashed) patients of four cortical connections selected to demonstrate our statistical framework. Empirical means are shown as vertical lines. All connectivities for healthy and schizophrenic patients have substantial overlap, evidenced by the darker regions in the histograms.

Our data space is therefore vastly reduced, from $[0, \infty[^{10^{17}}$ to $[-1, 1]^{40}$.

We let the symbol $\boldsymbol{f}$ stand for the set of connectivity weights extracted from the fMRI

data, rather than the full fMRI data themselves. With this new meaning of the symbol $\boldsymbol{f}$, the likelihoods for the health conditions (2.5) and de Finetti's formulae (2.7)–(2.8) remain formally unchanged, but now involve a data space with much fewer dimensions.

### 2.2.4 Trimming the distribution space: models by sufficiency and generalized normals

#### 2.2.4.1 Parametric models

The integrals in de Finetti's formulae (2.7)–(2.8) still represent a mixture of all imaginable pairs of probability distributions $(q_{\mathrm{H}}, q_{\mathrm{S}})$ over the space $[-1,1]^d$, and are therefore extremely complex. We now examine two ways to reduce this integral to a manageable set of distributions and to obtain analytically tractable formulae.

In the Bayesian literature, the complication of considering all possible distributions $q_c(\boldsymbol{f})$ of the quantity $\boldsymbol{f}$ is typically sidestepped by restricting them to a finite-dimensional set of distributions $L_c(\boldsymbol{f}|\,\boldsymbol{\theta}_c)$, identified or indexed by a finite number of parameters $\boldsymbol{\theta}_c$. For this reason, such a set is called a parametric family of distributions. An example of a parametric family is the set of $d$-variate normal distributions parameterized by their mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. With this restriction, the integrals in de Finetti's formulae (2.7) and (2.8) represent mixtures of distributions within the parametric family, the weight for each distribution being represented by a weight for its parameters. That is, we are no longer considering mixtures of all possible multivariate truncated normals, gammas, exponentials, etc., as in the example of Sec. 2.2.2, but only mixtures of truncated normals, say, with different means and covariance matrices. These integrals are thus ordinary finite-dimensional integrals. What happens to formulae (2.7)–(2.8) is that

$$q_c(\boldsymbol{f}) \text{ is replaced by } L_c(\boldsymbol{f}|\,\boldsymbol{\theta}_c), \qquad \mathrm{p}(q_{\mathrm{H}}, q_{\mathrm{S}}|\,I)\,\mathrm{d}q_{\mathrm{H}}\,\mathrm{d}q_{\mathrm{S}} \text{ is replaced by } \mathrm{p}(\boldsymbol{\theta}_{\mathrm{H}}, \boldsymbol{\theta}_{\mathrm{S}}|\,M, I)\,\mathrm{d}\boldsymbol{\theta}_{\mathrm{H}}\,\mathrm{d}\boldsymbol{\theta}_{\mathrm{S}}. \tag{2.10}$$

The distribution $L_c$ is called the likelihood of the parameters $\boldsymbol{\theta}_c$ , and $\mathrm{p}(\boldsymbol{\theta}_{\mathrm{H}}, \boldsymbol{\theta}_{\mathrm{S}}|\,M, I)$ is the prior parameter distribution. A parametric family and a prior distribution over its parameters are jointly called a parametric statistical model, which we denote by $M$. The term "model" is justly criticized by some probability theorists (see e.g. Besag & Kalman in Besag et al., 2002) but widely used, so we shall adopt it here.

In our present problem, the parametric statistical model needs not be the same for all health conditions: for example, we could use normal distributions for one condition and beta distributions for another, if that choice had better reflected the distributions of connectivity weights under the two different health conditions. For this reason, we use the subscript "$c$" in the formulae above. The likelihood we want to determine, Eq. (2.5), thus becomes, following

de Finetti's formula Eq. (2.8) and applying the replacements (2.10),

$$\mathrm{p}(F_0^{\boldsymbol{f}} \,|\, C_0^c \wedge D \wedge M \wedge I) = \int L_c(\boldsymbol{f}|\,\boldsymbol{\theta}_c)\,\mathrm{p}(\boldsymbol{\theta}_{\mathrm{H}}, \boldsymbol{\theta}_{\mathrm{S}}|\,D, M, I)\,\mathrm{d}\boldsymbol{\theta}_{\mathrm{H}}\,\mathrm{d}\boldsymbol{\theta}_{\mathrm{S}} \tag{2.11a}$$

$$\text{with}\quad \mathrm{p}(\boldsymbol{\theta}_{\mathrm{H}}, \boldsymbol{\theta}_{\mathrm{S}}|\,D, M, I) = \frac{\left[\prod_i L_{c_i}(\boldsymbol{f}_i|\,\boldsymbol{\theta}_{c_i})\right]\mathrm{p}(\boldsymbol{\theta}_{\mathrm{H}}, \boldsymbol{\theta}_{\mathrm{S}}|\,M, I)}{\int \left[\prod_i L_{c_i}(\boldsymbol{f}_i|\,\boldsymbol{\theta}_{c_i})\right]\mathrm{p}(\boldsymbol{\theta}_{\mathrm{H}}, \boldsymbol{\theta}_{\mathrm{S}}|\,M, I)\,\mathrm{d}\boldsymbol{\theta}_{\mathrm{H}}\,\mathrm{d}\boldsymbol{\theta}_{\mathrm{S}}}. \tag{2.11b}$$

#### 2.2.4.2 Models by sufficient statistics

The choice of a statistical model often appears as an art and a matter of experience. Notable statisticians have voiced concerns over the esoteric character of this choice. Dawid (1982a, p. 220) says: "Where do probability models come from? To judge by the resounding silence over this question on the part of most statisticians, it seems highly embarrassing". And Diaconis (1988, Sec. 8, p. 121) remarks:

> de Finetti's alarm at statisticians introducing reams of unobservable parameters has been repeatedly justified in the modern curve fitting exercises of today's big models. These seem to lose all contact with scientific reality focusing attention on details of large programs and fitting instead of observation and understanding of basic mechanism. It is to be hoped that a fresh implementation of de Finetti's program based on observables will lead us out of this mess.

Authors like these have also tried to develop intuitive methods for choosing a model, for example by proving that a parametric family can be uniquely determined by some symmetry assumptions about our inferences, or by other information-theoretical properties (Bernardo & Smith, 2000, Ch. 4; Lindley, 2008, Sec. 5.5; an enlightening discussion of this topic is given by Dawid, 2013). In the present study we want to emphasize, as Cifarelli & Regazzini (1982) did, that a statistical model can be chosen by selecting a *sufficient statistics* (Kolmogorov, 1942; Freedman, 1962; Diaconis & Freedman, 1980, 1981; Cifarelli & Regazzini, 1982; Lauritzen, 1988; Diaconis, 1992; Kallenberg, 2005, and the textbook references above). Here is an example.

Imagine that we have patients labelled $i' \in \{1', 2', \dots\}$, and $n$ patients labelled $i \in \{1, 2, \dots\}$, all with the same health condition. Of the second set of patients we also know the connectivity weights $\{\boldsymbol{f}_i\}$ obtained by fMRI. We want to specify the joint probability distribution $\mathrm{p}(\{\boldsymbol{f}_{i'}\}|\,\{\boldsymbol{f}_i\}, I)$ that the fMRIs of the patients $\{i'\}$ yield connectivity weights $\{\boldsymbol{f}_{i'}\}$, conditional on our knowledge of the connectivity weights of the $n$ patients $\{i\}$. Now assume that the probabilities for the fMRI results are exchangeable, so that de Finetti's formulae (2.7) and (2.8) hold. Also assume that in order to specify the joint distribution we do not need the full set of data $\{\boldsymbol{f}_i\}$, but only their number $n$ and some statistics, e.g. the empirical mean and covariance matrix of these data,

$$\overline{\boldsymbol{f}} \coloneqq \frac{1}{n}\sum_i \boldsymbol{f}_i, \qquad \mathrm{Cov}(\boldsymbol{f}) \coloneqq \frac{1}{n}\sum_i (\boldsymbol{f}_i - \overline{\boldsymbol{f}})(\boldsymbol{f}_i - \overline{\boldsymbol{f}})^{\intercal}; \tag{2.12}$$

the rest of the details of the data $\{\boldsymbol{f}_i\}$ being irrelevant. In other words, we are assuming that the statistics above are *sufficient* for us to make predictions as if we had the full data. In

symbols,

$$\mathrm{p}(\{\boldsymbol{f}_{i'}\}|\,\{\boldsymbol{f}_i\}, I) = \mathrm{p}(\{\boldsymbol{f}_{i'}\}|\,n, \overline{\boldsymbol{f}}, \mathrm{Cov}\,\boldsymbol{f}, I). \tag{2.13}$$

If this is true regardless of the number of patients $\{i'\}$ and $\{i\}$, then these statistics are called *(predictive) sufficient statistics*. There are several notions of sufficiency, including the traditional one by Fisher (1922) and Neyman (1935), but they all are more or less equivalent (Bernardo & Smith, 2000, Sec. 4.5.2).

The assumption of the existence of sufficient statistics has a very important consequence for de Finetti's formulae (2.7) and (2.8): the space of possible prior distributions is hugely reduced, constrained to be non-zero only over a parametric family of distributions that is determined by the sufficient statistics. The replacement (2.10) takes place, leading to the simpler formula (2.11) for the likelihoods of the health conditions. The number of parameters is equal to that of the sufficient statistics. The proof of this reduction was given by Pitman and Koopman (Koopman, 1936; Pitman, 1936; Darmois, 1935; for generalizations, e.g. to the discrete case, see Hipp, 1974; Andersen, 1970; Denny, 1967; Fraser, 1963; Barankin & Maitra, 1963). When the sufficient statistics are the mean and covariance matrix, as above, the likelihoods turn out to be (truncated) multivariate normal distributions.

In the slightly more complicated case of two or more health conditions and the assumption of partial exchangeability (2.6), this theorem leads to formula (2.11) with likelihoods $L_c$ determined by the sufficient statistics that we have chosen for the different health conditions (Bernardo & Smith, 2000, Sec. 4.6). The prior distribution over the parameters of the likelihoods, $\mathrm{p}(\boldsymbol{\theta}_{\mathrm{H}}, \boldsymbol{\theta}_{\mathrm{S}}|\, I)$, is not determined by the theorem, but has to be determined by other considerations that can again involve symmetry and information theory.

As we mentioned in the introduction, the notion of sufficient statistics can be a helpful bridge between biophysical considerations and the specification of probabilities. It may be easier to conceive and understand a connection between biophysical considerations and some statistics of our measurements, than between biophysical considerations and an abstract multidimensional distribution function. Once such statistics are selected, they in turn uniquely select a probability distribution for us. Vice versa, if a statistical model based on some sufficient statistics proves to be very reliable in its predictions, we may conclude that its sufficient statistics must have an important biological meaning.

In the remaining study we shall use three statistical models determined by three different sufficient statistics. Our choice of statistics is unfortunately not biologically motivated, as such models have yet to be determined for fMRI data. However, they are adequate to demonstrate the approach and we hope that authors with more experience will pursue this line of thought and derive better-motivated sufficient statistics.

#### 2.2.4.3 Edgeworth's "method of translation": generalized normal models

The assumption of a sufficient statistics makes the integrals in de Finetti's formulae (2.7) and (2.8) finite-dimensional, but these integrals and other expressions that depend on them, like the post-test probabilities, may still lack a closed form and be analytically intractable. In

this case we could use numerical methods, a computationally costly possibility we consider in the Discussion, Sec. 2.3.3. In the present work we choose models with closed-form formulae instead; their swift calculation facilitates the model comparison to be illustrated later.

Our starting point is an analytically tractable model by sufficient statistics that has been the subject of much study (Gelman et al., 2014, Sec. 3.6; Minka, 2001; Murphy, 2007): it has a normal likelihood, with mean $\boldsymbol{\lambda}$ and covariance matrix $\boldsymbol{\Lambda}$ as parameters, and a normal-inverse-Wishart prior distribution over these parameters. This parameter prior maintains the same functional form when updated with training data; this kind of prior is called *conjugate* (DeGroot, 2004, Ch. 9; Diaconis & Ylvisaker, 1979). This model is outlined in more detail in Sec. 2.4.2.

We try to make the normal + normal-inverse-Wishart model more flexible by combining it with an idea that Edgeworth (1898) called "Method of Translation", discussed also by Johnson (1949) and Mead (1965): the transformation of a quantity into a normally distributed one. That is, instead of considering the connectivity weights $\boldsymbol{f}$, we consider transformed quantities $l(\boldsymbol{f})$, where $l$ is a component-wise monotonic function, and suppose the latter quantities to be normally distributed. This leads to generalized-normal likelihoods of the form

$$\mathrm{N}[l(\boldsymbol{f})|\,\boldsymbol{\lambda}, \boldsymbol{\Lambda}]\; l'(\boldsymbol{f})\,\mathrm{d}\boldsymbol{f} \tag{2.14}$$

where N is the normal distribution with mean $\boldsymbol{\lambda}$ and covariance matrix $\boldsymbol{\Lambda}$, and $l'$ is the Jacobian determinant of $l$.

This simple idea has an amazing scope: it allows us to explore a vast range of non-normal likelihoods – in particular likelihoods defined on bounded domains such as $[-1,1]^d$ – and to keep the low computational costs of the conjugate prior. In our present problem it has also an additional convenient feature: in the calculation of the post-test probability, the Jacobian determinant $l'$ disappears, as Eq. (2.19) below shows.

These generalized-normal models, which we denote by $M_l$, are also determined by a choice of sufficient statistics, analogous to Eq. (2.12): the mean and covariance matrix of the transformed data $\{l(\boldsymbol{f}_i)\}$,

$$\overline{l(\boldsymbol{f})}, \qquad \mathrm{Cov}[l(\boldsymbol{f})]. \tag{2.15}$$

In our partially exchangeable case, with formulae (2.11), we need to specify a likelihood $L_c$ for each health condition $c$. We assume these two likelihoods $L_\mathrm{H}$, $L_\mathrm{S}$ to be functionally identical generalized normals, i.e. the function $l$ is the same for the healthy and the schizophrenic case; but their means and covariance matrices $(\boldsymbol{\lambda}_\mathrm{H}, \boldsymbol{\Lambda}_\mathrm{H})$ and $(\boldsymbol{\lambda}_\mathrm{S}, \boldsymbol{\Lambda}_\mathrm{S})$ can be different.

We also need to specify a joint parameter prior for $(\boldsymbol{\lambda}_\mathrm{H}, \boldsymbol{\Lambda}_\mathrm{H}; \boldsymbol{\lambda}_\mathrm{S}, \boldsymbol{\Lambda}_\mathrm{S})$. To use the analytic advantage of the conjugate prior, we assume that the distribution for these parameters is a product of independent distributions:

$$\mathrm{p}(\boldsymbol{\lambda}_\mathrm{H}, \boldsymbol{\Lambda}_\mathrm{H}; \boldsymbol{\lambda}_\mathrm{S}, \boldsymbol{\Lambda}_\mathrm{S}|\, M_l, I) = \mathrm{p}(\boldsymbol{\lambda}_\mathrm{H}, \boldsymbol{\Lambda}_\mathrm{H}|\, M_l, I) \times \mathrm{p}(\boldsymbol{\lambda}_\mathrm{S}, \boldsymbol{\Lambda}_\mathrm{S}|\, M_l, I), \tag{2.16}$$

each of them being a normal-inverse-Wishart distribution described in Sec. 2.4.2. This

independence assumption is quite strong and has an important consequence: *the likelihood for a health condition only depends on the data from previous patients having that same health condition*.

With the assumptions above, the likelihood for the health condition needed by the clinician has a closed form for this model (see Sec. 2.4.2). The likelihood for patient $0$'s being healthy, in view of the patient's measured connectivity weights $\boldsymbol{f}$, and given the data $(\boldsymbol{f}_i)$ from previous $n_{\mathrm{H}}$ healthy patients, is

$$\mathrm{p}(\text{fMRI result } \boldsymbol{f} \mid \text{ healthy } \wedge \text{ prior info}) = \mathrm{t}\Big[l(\boldsymbol{f})\Big|\ \nu_{\mathrm{H}} - d + 1, \boldsymbol{\delta}_{\mathrm{H}}, \tfrac{\kappa_{\mathrm{H}}+1}{\kappa_{\mathrm{H}}\,(\nu_{\mathrm{H}}-d+1)}\boldsymbol{\Delta}_{\mathrm{H}}\Big] \prod_k l'(f_k), \tag{2.17}$$

where $\mathrm{t}$ is a multivariate t distribution with $\nu_{\mathrm{H}} - d + 1$ degrees of freedom, mean $\boldsymbol{\delta}_{\mathrm{H}}$, and scale matrix $\frac{\kappa_{\mathrm{H}}+1}{\kappa_{\mathrm{H}}\,(\nu_{\mathrm{H}}-d+1)}\boldsymbol{\Delta}_{\mathrm{H}}$. This distribution has covariance matrix $\frac{\kappa_{\mathrm{H}}+1}{\kappa_{\mathrm{H}}\,(\nu_{\mathrm{H}}-d-1)}\boldsymbol{\Delta}_{\mathrm{H}}$ (note the different denominator from the scale matrix), and approaches a generalized normal for large $\nu_{\mathrm{H}}$. The final factor is the Jacobian determinant of $l$.

The most important feature of this likelihood is the dependence of the coefficients $(\kappa_{\mathrm{H}}, \boldsymbol{\delta}_{\mathrm{H}}, \nu_{\mathrm{H}}, \boldsymbol{\Delta}_{\mathrm{H}})$ on the data $(\boldsymbol{f}_i)$ of the previous $n_{\mathrm{H}}$ healthy patients:

$$\begin{aligned}
&\kappa_{\mathrm{H}} = \kappa_0 + n_{\mathrm{H}}, && \nu_{\mathrm{H}} = \nu_0 + n_{\mathrm{H}},\\
&\boldsymbol{\delta}_{\mathrm{H}} = \frac{\kappa_0\,\boldsymbol{\delta}_0 + n_{\mathrm{H}}\,\overline{l(\boldsymbol{f})}}{\kappa_0 + n_{\mathrm{H}}}, && \boldsymbol{\Delta}_{\mathrm{H}} = \boldsymbol{\Delta}_0 + n_{\mathrm{H}}\ \mathrm{Cov}[l(\boldsymbol{f})] + \frac{\kappa_0\, n_{\mathrm{H}}}{\kappa_0 + n_{\mathrm{H}}}\big[\overline{l(\boldsymbol{f})} - \boldsymbol{\delta}_0\big]\big[\overline{l(\boldsymbol{f})} - \boldsymbol{\delta}_0\big]^{\intercal},
\end{aligned} \tag{2.18}$$

where $(\kappa_0, \boldsymbol{\delta}_0, \nu_0, \boldsymbol{\Delta}_0)$ are prior coefficients that represent the clinician's knowledge before any patients were observed. As the number $n_{\mathrm{H}}$ of observed healthy patients increases, the probability for the transformed data $l(\boldsymbol{f})$ tends to a normal distribution with mean and covariance matrix equal to the empirical average and covariance matrix of the transformed data. The formulae above show that previous data enter only through the sufficient statistics $\overline{l(\boldsymbol{f})}$ and $\mathrm{Cov}[l(\boldsymbol{f})]$.

An analogous formula holds for the likelihood for the patient's being schizophrenic, with coefficients $(\kappa_{\mathrm{S}}, \boldsymbol{\delta}_{\mathrm{S}}, \nu_{\mathrm{S}}, \boldsymbol{\Delta}_{\mathrm{S}})$ that depend on the data of previous schizophrenic patients and some initial coefficients. The function $l$ and the prior coefficients $(\kappa_0, \boldsymbol{\delta}_0, \nu_0, \boldsymbol{\Delta}_0)$ could be different for the healthy and schizophrenic cases, but for simplicity we assume them identical for both health conditions.

If $(\hat{p}_{\mathrm{H}}, \hat{p}_{\mathrm{S}})$ is the pre-test probability distribution for the health condition of patient $0$, his post-test probability to be healthy is

$$\mathrm{P}(\text{healthy} \mid \text{fMRI result} \wedge \text{prior info}) = \frac{\mathrm{t}\Big[l(\boldsymbol{f})\Big|\ \nu_\mathrm{H} - d + 1, \boldsymbol{\delta}_\mathrm{H}, \frac{\kappa_\mathrm{H}+1}{\kappa_\mathrm{H}\,(\nu_\mathrm{H}-d+1)}\boldsymbol{\Delta}_\mathrm{H}\Big]\,\hat{p}_\mathrm{H}}{\mathrm{t}\Big[l(\boldsymbol{f})\Big|\ \nu_\mathrm{H} - d + 1, \boldsymbol{\delta}_\mathrm{H}, \frac{\kappa_\mathrm{H}+1}{\kappa_\mathrm{H}\,(\nu_\mathrm{H}-d+1)}\boldsymbol{\Delta}_\mathrm{H}\Big]\,\hat{p}_\mathrm{H} + \mathrm{t}\Big[l(\boldsymbol{f})\Big|\ \nu_\mathrm{S} - d + 1, \boldsymbol{\delta}_\mathrm{S}, \frac{\kappa_\mathrm{S}+1}{\kappa_\mathrm{S}\,(\nu_\mathrm{S}-d+1)}\boldsymbol{\Delta}_\mathrm{S}\Big]\,\hat{p}_\mathrm{S}}. \tag{2.19}$$

Note that the Jacobian determinants $l'$ do not appear in this formula.

#### 2.2.4.4 Generalized normal models in our study

In the rest of our study we compare three different transformations $l$ of the connectivity weights $\boldsymbol{f}$, with one set of prior coefficients $(\kappa_0, \boldsymbol{\delta}_0, \nu_0, \boldsymbol{\Delta}_0)$ each:

**Logit-normal model:** A slightly modified logit transformation

$$l(f_k) \coloneqq \ln \frac{1+f_k}{1-f_k}, \qquad l'(f_k) = \frac{2}{1-f_k^2}, \tag{2.20}$$

with prior coefficients

$$\kappa_0 = 1, \quad \boldsymbol{\delta}_0 = 0, \quad \nu_0 = d+1, \quad \boldsymbol{\Delta}_0 = (d+2)\boldsymbol{I}_d. \tag{2.21}$$

**Tangent-normal model:** A tangent transformation

$$l(f_k) \coloneqq \tan \frac{\pi f_k}{2}, \qquad l'(f_k) = \frac{\pi}{1+\cos \pi f_k}, \tag{2.22}$$

with prior coefficients

$$\kappa_0 = 1, \quad \boldsymbol{\delta}_0 = 0, \quad \nu_0 = d+1, \quad \boldsymbol{\Delta}_0 = \frac{d+2}{4}\boldsymbol{I}_d. \tag{2.23}$$

**Normal model:** An identity transformation (that is, no transformation at all)

$$l(f_k) \coloneqq f_k, \qquad l'(f_k) = 1, \tag{2.24}$$

with prior coefficients

$$\kappa_0 = 1, \quad \boldsymbol{\delta}_0 = 0, \quad \nu_0 = d+1, \quad \boldsymbol{\Delta}_0 = 10\boldsymbol{I}_d. \tag{2.25}$$

For brevity we shall denote $l(\boldsymbol{f}) \coloneqq \big(l(f_i)\big)$.

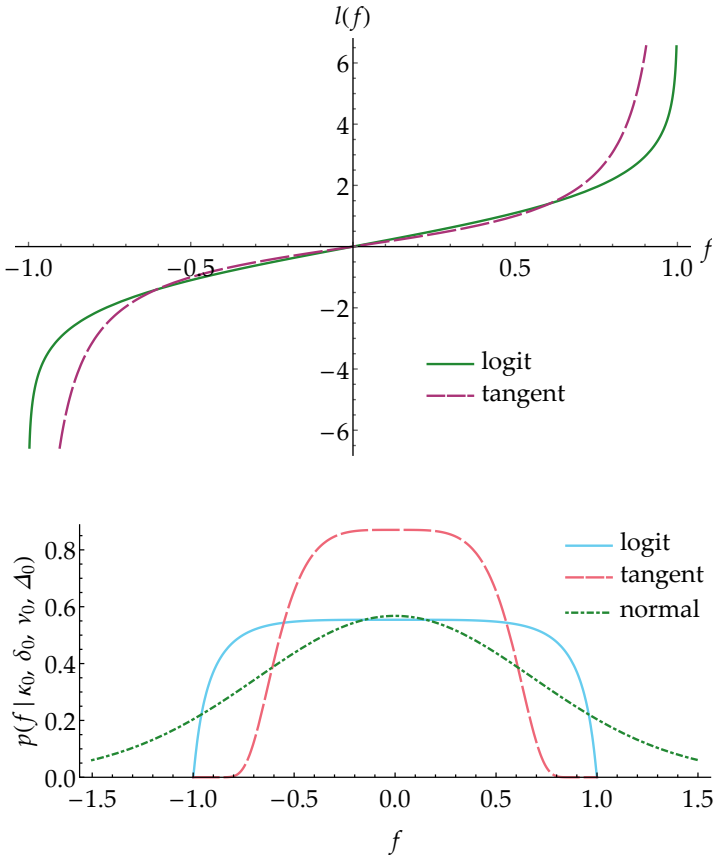


**Figure 2.2:** General normalized models. Upper panel: Generalized logit and tangent transformation functions as defined in the text. Lower panel: Prior distributions for any connectivity weight $f$ conditional on the three generalized normal models defined in the text

The first two transformations, plotted in the upper panel of Fig. 2.2, map the bounded domain $]-1,1[$ of the connectivity weights into the reals, and thus restrict the generalized-normal likelihood to meaningful values of the connectivity weights. The last model instead allows for connectivity weights outside their meaningful bounds. It can be conceived as the model of a person who has no precise knowledge of what the quantities $\boldsymbol{f}$ are. Since the clinician's final predictions concern health conditions given data $\boldsymbol{f}$, not the data $\boldsymbol{f}$ themselves, this model can still be meaningfully used. The probabilities for the connectivity weight $f_i$ conditional on the prior coefficients above are shown in the lower panel of Fig. 2.2.

The prior coefficients are chosen by the following criteria: uniform marginal distributions of the correlations between connectivity weights (leading to $\nu_0 = d+1$ as explained before); large uncertainty in the location parameters ($\kappa_0 = 1$); symmetry with respect to the origin ($\boldsymbol{\delta}_0 = 0$); a prior distribution for the connectivity as flat as possible (its second derivative vanishes

at the origin, leading to the values of $\boldsymbol{\Delta}_0$ above). In the case of the identity transformation we have chosen a $\boldsymbol{\Delta}_0$ that somewhat concentrates the prior around the true range of the connectivity weights, $[-1, 1]$.

The numerical values of the main quantities used throughout this study are summarized in table 2.1.

| | |
|---|---|
| $n_{\mathrm{H}} = 55$ | healthy patients |
| $n_{\mathrm{S}} = 49$ | schizophrenic patients |
| $d = 40$ | graph parameters |
| $\kappa_0, \boldsymbol{\delta}_0, \nu_0, \boldsymbol{\Delta}_0$ | prior coefficients: see (2.23), (2.25) |

**Table 2.1:** Numerical values in our study

### 2.2.5 Model comparison and selection

#### 2.2.5.1 Criteria for model comparison

Any two statistical models differ in two main characteristics: their predictive power and their learning speed. Predictive power is a model's capacity to give high probability to propositions that turn out to be true, during and especially after its learning phase (cf. Dawid, 1982b). Learning speed is how quickly a model reaches unchanging, stable predictive probabilities as it gets updated with new data; note, however, that a model may also never reach stable probabilities (see e.g. Bruno, 1964; Berk, 1966); "Alas, this seems like a model of the way things work in practical inference – as more data comes in, one admits a richer and richer variety of explanatory hypothesis" (Diaconis, 1988, Sec. 3, p. 113). Thus, our choice of a model depends on the relative importance we give to these two characteristics.

These two characteristics need not go hand in hand: a model can quickly learn with very little data but settle on probabilities with poor predictive power; conversely, it can reach great predictive power but only after a long learning phase with a huge amount of data. A model can also "unlearn", i.e. its predictive power initially increases and then decreases before stabilizing. The latter phenomenon can happen because every model initially makes its prediction through a mixture of likelihoods – the integral (2.11), in our case the t distribution (2.36) – but as the training continues the mixture is replaced by a single likelihood, in our case a generalized normal (2.14), as explained in Sec. 2.2.4.3. It may happen that the mixture of likelihoods has higher predictive power than a single likelihood, and in this case a decline in predictive power will be observed. A model with such behaviour is obviously unfit for the clinician's goal; this also means that the sufficient statistics on which it is based capture very poorly the differences in connectivity weights between health conditions.

When only a small amount of training data is available, it can be difficult to assess which of two models has or will have the greater predictive power. The first model can initially

reach a greater predictive power than the second, but the second model may eventually reach greater predictive power than the first, with further training data. Models having the same likelihood, however, have the same final predictive power; their learning speeds depend on their parameter priors.

In a diagnostic problem like that faced by our clinician, the choice of a model is ultimately dictated by the predictive power of the post-test probabilities given by the model; but if we have little training data, the learning speed of the model is also of some importance. Several quantitative criteria can be conceived to assess these two characteristics:

1. the post-test probabilities the model gives to the correct health conditions for all training data, i.e. $\mathrm{P}(c_1, c_2, \dots)$;
2. the post-test probabilities the model gives to the correct health conditions in the final phase of the training only, i.e. $\mathrm{P}(c_{\text{last}} | \, c_1, c_2, \dots)$;
3. the expected utility the model yields for all training data;
4. the expected utility the model yields in the final phase of the training only.

Post-test probabilities (criteria 1, 2) are important for obvious reasons, and utilities (criteria 3, 4) are important because the clinician's overall problem is one of decision, as emphasized in the Introduction of this chapter. Consideration of all data (criteria 1, 3) is important if we are interested in the performance of the model for the whole set of patients; but consideration of the final data only, conditional on the previous ones (criteria 2, 4), tells us how much the model has learned (compare with a similar remark in model comparison using Bayes factors by Berger & Pericchi, 1996).

We must keep in mind that these criteria assess a statistical model not by itself but in combination with other factors, since they also depend on pre-test probabilities (which can be influenced by other diagnostic tests) or utilities.

Applied to our models, each of these four criteria gives a very similar picture. We shall calculate the results for criteria 2 and 4, the latter with two different utility tables. This calculation can be explained in very intuitive terms:

Imagine that the $n_{\mathrm{H}} = 55$ healthy and $n_{\mathrm{S}} = 49$ schizophrenic patients visit the clinician in turn, in an unknown order. For each patient, let us further assume that the clinician has pre-test probabilities $(\hat{p}_{\mathrm{H}}, \hat{p}_{\mathrm{S}}) = (0.5, 0.5)$, i.e. she is completely uncertain about the patient's health condition. The incidence in the population is much lower than 50%, of course, but the patients presenting themselves for diagnosis are not representative of the full population.

As stated in Sec. 2.2.2, pre-test probabilities represent the clinician's uncertainty before the fMRI test is made; they can be based for example on a first diagnosis considering symptoms and medical history of the patient and of the patient's family, on psychological evaluations, and on other diagnostic tests. Here we assume complete uncertainty for demonstration purposes.

The clinician acquires the fMRI result $\boldsymbol{f}$ for that patient, and uses the statistical model, trained with the data from all the patients that previously visited her, to update to a post-test probability for schizophrenia $p_{\mathrm{S}}$, given by Eq. (2.19); obviously $p_{\mathrm{H}} = 1 - p_{\mathrm{S}}$. Now the clinician must make a decision – say, treat or dismiss – based on the expected utilities of the decisions available. Each decision has a different utility depending on the patient's true health condition, as summarized by a table. We consider two tables: a symmetric one

| (symmetric) | healthy | schizophrenic |
|---|---|---|
| dismiss | 1 | 0 |
| treat | 0 | 1 |

(2.26)

and an asymmetric one

| (asymmetric) | healthy | schizophrenic |
|---|---|---|
| dismiss | 1 | −2 |
| treat | −1 | 2 |

(2.27)

In order to maximize the expected utility, as explained in Sec. 2.4.3, the clinician dismisses the patient if $p_{\mathrm{S}} < 1/2$ in the case of the symmetric utility table, and if $p_{\mathrm{S}} < 1/3$ in the case of the asymmetric one; and treats the patient otherwise. After the clinician's decision is made, the patient's true health condition is revealed and we record the actual utility gained for each table, and the post-test probability the clinician assigned to the true health condition, or its logarithm, usually called "negative surprise" (Bartlett, 1952; Good, 1956, 1957a). The health condition and fMRI data of this patient are used to update the model. The next patient is received, and so on, until all patients have been examined.

The particular sequence of utilities and log-probabilities recorded in the manner just described depends on the exact order by which the 104 patients visit the clinician. We take an approximate average over all possible $104! \approx 10^{166}$ orders by randomly sampling $520\,000$ of them. The values of these averages for the final patient constitute the quantitative criteria 2 and 4.

With the symmetric utility table (2.26), the average utility is also the average number of schizophrenic patients for which the model yields $p_{\mathrm{S}} > 1/2$. The average utility for the last patient, when the model has been trained with the rest of the patients, is therefore a form of leave-one-out cross-validation (Allen, 1974; Stone, 1974; Kotz et al., 2006, vol. 2, pp. 1454–1458). The asymmetric table (2.27), slightly more realistic, tells us that dismissing a schizophrenic patient has worse consequences than treating a healthy one, and treating a schizophrenic patient has better consequences than dismissing a healthy one (McKenzie, 2014; Ho et al., 2000). For this reason the patient is dismissed, more conservatively, only if $p_{\mathrm{S}} < 1/3$. Note that scaling a utility table by a positive factor or shifting its values by a constant represent changes in the unit of measure and in the zero of utilities, and therefore do not affect our relative comparison of the statistical models.

#### 2.2.5.2 Results for our three models

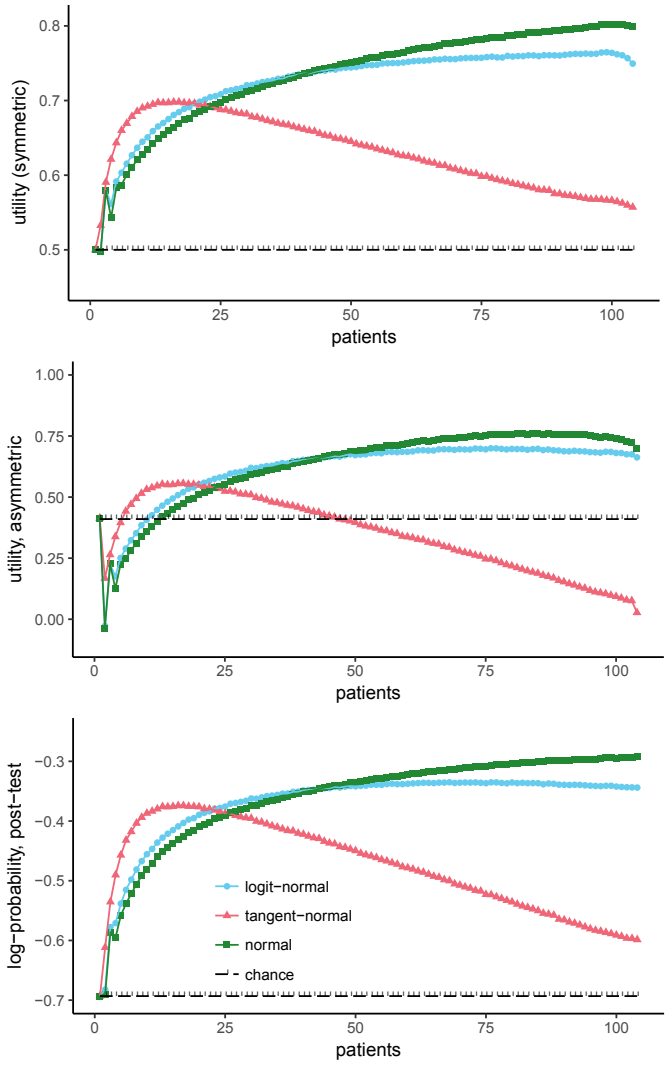


**Figure 2.3:** Averaged sequence of utilities, with utility tables (2.26) and (2.27), and of log-probabilities for our three models. The standard deviations of the averages are smaller than the markers' size. The average values for the first and last patients are exact.

The averaged sequences of utilities and log-probabilities calculated as in Sec. 2.2.5.1 are shown in Fig. 2.3. The R code for the calculation is publicly available (Porta Mana et al., 2018). The results, summarized in table 2.2, are qualitatively identical by the two criteria and two utility tables we chose: the normal model gives the best values for the final patient, followed

by the logit-normal model; the tangent-normal model has the worst final predictive power, at or below chance level.

| | symmetric utility (2.26) | asymmetric utility (2.27) | log-probability |
|---|---|---|---|
| normal model | $0.80$ | $0.70$ | $-0.29$ |
| logit-normal model | $0.75$ | $0.66$ | $-0.34$ |
| tangent-normal model | $0.56$ | $0.03$ | $-0.60$ |
| chance | $0.50$ | $0.41$ | $-0.69$ |

**Table 2.2:** Final results for the three models

The plots of Fig. 2.3 illustrate the points made at the beginning of Sec. 2.2.5.1: one model can initially learn faster than another and yet be overtaken in the later stages of learning; this is the case for the logit-normal and normal models. The tangent-normal model shows strong unlearning; this means that a tangent-normal likelihood and its sufficient statistics do not distinguish well between healthy and schizophrenic conditions. The slight downward bends at the final stages of the logit-normal and normal models raise the suspicion that they might also show some unlearning if further training data were supplied.

The trends of the logit-normal and normal models suggest that the learning phase is not finished: more patients are needed before their predictive probabilities become stable. This is also evident from the updated marginal distributions of their parameters $(\boldsymbol{\lambda}_\mathrm{H}, \boldsymbol{\Lambda}_\mathrm{H}; \boldsymbol{\lambda}_\mathrm{S}, \boldsymbol{\Lambda}_\mathrm{S})$, for example those for the connectivity weight $f$ between the left superior parietal lobule and left lingual gyrus, shown in Fig. 2.4 for the logit-normal model. The distributions of the location parameters $(\boldsymbol{\lambda}_\mathrm{H}, \boldsymbol{\lambda}_\mathrm{S})$ have reached the empirical means of the data, but those of the scale parameters $(\boldsymbol{\Lambda}_\mathrm{H}, \boldsymbol{\Lambda}_\mathrm{S})$ are still very far away from the empirical variances. The reason is that the prior for the scale parameters had a peak at a very large value of $\Lambda \approx 20$. The $55$ data for healthy patients and $49$ for schizophrenic ones have shifted this peak to $\Lambda_\mathrm{H} = 1.3$ and $\Lambda_\mathrm{S} = 1.5$, but more data are needed to shift these peaks to even smaller values – provided that in the meantime the empirical values do not change too much as new data are gathered.

The peak at high values of $\boldsymbol{\Lambda}$ is a known inconvenient feature of the normal-inverse-Wishart conjugate prior, related to the correlation between correlation and variance components of $\boldsymbol{\Lambda}$ characteristic of this prior (e.g. Barnard et al., 2000).

#### 2.2.5.3 Contrast with other model-comparison criteria

Common Bayesian model-comparison criteria are based on the joint probability that the model gives to training data; especially its logarithm, called "weight of evidence" or "marginal log-likelihood", or the ratio of such logarithms, called Bayes factors (Jeffreys, 2003, Chs. V, VI, A; Good, 1950; MacKay, 1992a; Kass & Raftery, 1995). The simple reason is Bayes'

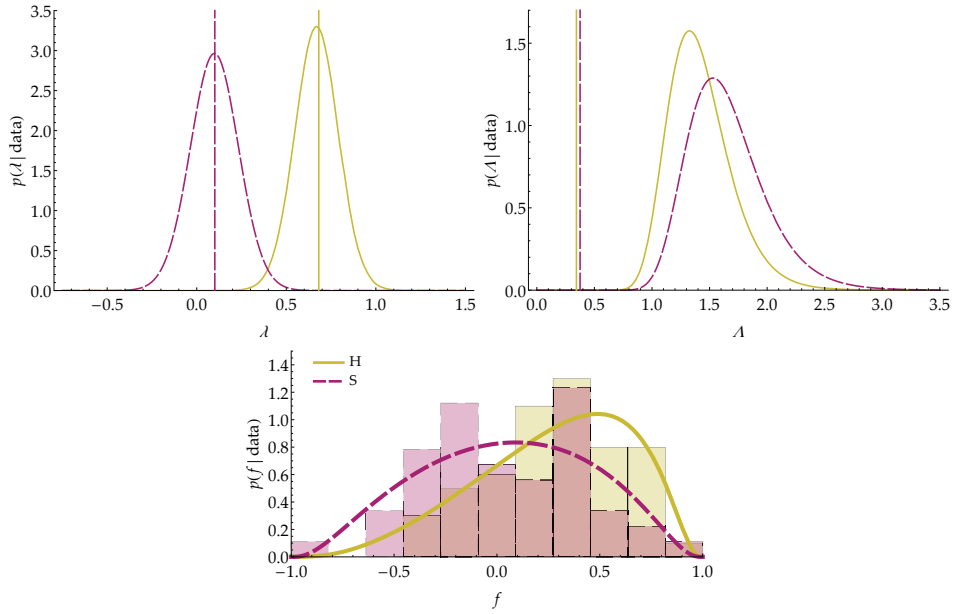


**Figure 2.4:** Updated distributions of the logit-normal model for the location parameters $(\boldsymbol{\lambda}_{\mathrm{H}}, \boldsymbol{\lambda}_{\mathrm{S}})$ (left), scale parameters $(\boldsymbol{\Lambda}_{\mathrm{H}}, \boldsymbol{\Lambda}_{\mathrm{S}})$ (right), and connectivity weights $f_{\mathrm{H}}$, $f_{\mathrm{S}}$ (bottom, superposed on the empirical distributions) for the connectivity between left superior parietal lobule and left lingual gyrus, corresponding to the bottom left panel of Fig. 2.1. The vertical lines in the first two plots indicate the corresponding empirical statistics from the data.

theorem:

$$\mathrm{P}(\text{model} \mid \text{data} \wedge \text{prior info}) \propto \mathrm{P}(\text{data} \mid \text{model} \wedge \text{prior info}) \times \mathrm{P}(\text{model} \mid \text{prior info}), \quad (2.28)$$

the latter probability usually assumed the same for all models (but see Porta Mana, 2017). A higher weight of evidence means that the model is more probable.

In our study, however, we have two kinds of data: health conditions and fMRI results. Since the likelihood for the health condition, used by the clinician to arrive at a post-test probability, gives the probability for the fMRI results $\{\boldsymbol{f}_i\}$, it seems intuitive to calculate the weights of evidence of our models based on these data. The result is the opposite of what we obtain with the averaged-utility criterion or any of other three mentioned above. We obtain:

$$\begin{aligned}
&\ln \mathrm{p}(\text{fMRI results} \mid \text{logit-normal model} \wedge \text{health conditions}) = -1913,\\
&\ln \mathrm{p}(\text{fMRI results} \mid \text{tangent-normal model} \wedge \text{health conditions}) = -1858,\\
&\ln \mathrm{p}(\text{fMRI results} \mid \text{normal model} \wedge \text{health conditions}) = -2488,
\end{aligned} \quad (2.29)$$

which gives the normal model a much smaller probability than the other two, and the tangent-normal model the highest.

This discrepancy with the averaged-utility criterion is not completely surprising, though. Imagine a disease that leads to no differences at all between the fMRI results of patients with the disease and those of healthy controls. If we found a statistical model that predicted the fMRI results with certainty, this model would thus have a the highest weight of evidence (zero), and yet its final average utility would be at chance level, since it could not help us at all in telling healthy from diseased patients. For our problem the right comparison and selection criterion is the utility or one of the other three criteria previously listed.

#### 2.2.5.4 Final assessment of models

The average-utility criterion clearly excludes the tangent-normal model, which even shows a rapid unlearning. The logit-normal and normal models have almost similar performances, at around 75–80% of final patients correctly treated. We can also plot the sequence of utilities averaged over healthy and schizophrenic patients separately, as in Fig. 2.5, which gives us an idea of the ratio between true and false negatives, and true and false positives. Both models show around 35% false positives (dismissed schizophrenic patients), with the normal model giving slightly higher final rates of true negatives, 93%, and true positives, 65%, than the normal, 85% and 63%.

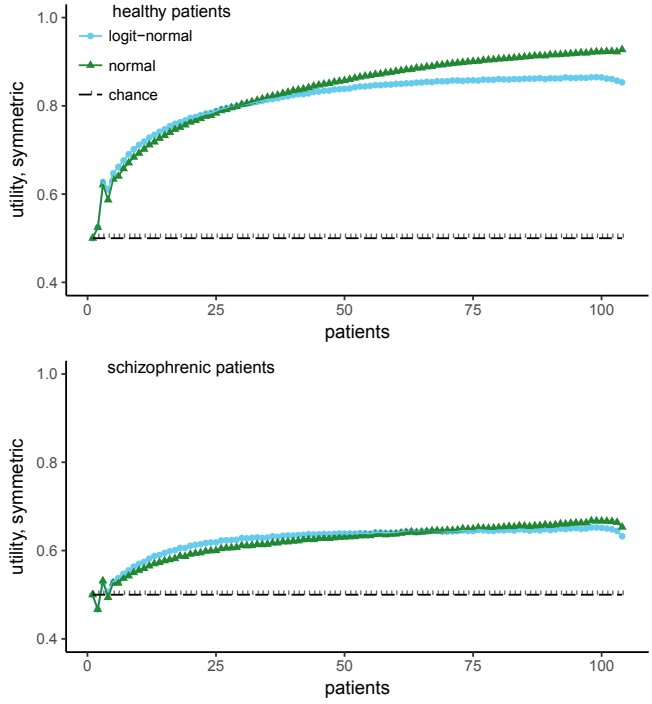


**Figure 2.5:** Sequence utilities averaged over healthy (left) and schizophrenic (right) patients separately, for the logit-normal and normal models

However, we emphasize that the features used as inputs to the model were selected using a very simple heuristic (maximum difference in means between the two groups, see Sec. 2.2.3), and the statistical model was selected for its computational properties rather than its fit to the distributions of connection weights derived from fMRI data. This notwithstanding, we believe that with further training, the logit-normal model could reach a higher predictive power. The reason is that some empirical distributions of connectivity weights, like the one for schizophrenic patients shown in red in Fig. 2.4, seem to be bimodal; and the logit-normal likelihood, unlike the normal one, is capable of bimodality, thus fitting these distributions better.

Our assessment, however, is just an illustrative example for the general method discussed in this work, and we are not earnestly proposing the logit-normal model (nor any other specific model) as the optimal one to use in the problem of diagnosing schizophrenia. We note that any model assessment and selection using the average-utility metric depends on several important quantities and assumptions:

A. the pre-test probabilities given by the clinician; we assumed these to be $(0.5, 0.5)$;

B. the clinician's range of decisions; we assumed it to be simply "treat or dismiss", but it could comprise several different kinds of treatments;

C. the utilities of the clinician's decisions; we assumed these to be as in formula (2.37);

D. the ratio between the numbers of healthy and schizophrenic training data; $55/49 = 1.12$ in our case;

E. the clinician diagnoses one patient at a time.

For a proper model assessment we should therefore investigate and consider more realistic rates of healthy vs schizophrenic cases that visit a particular clinician, in order to have better-informed pre-test probabilities; and we should consider more realistic decisions available to the clinician, together with realistically quantified utilities.

Assumption E. deserves some explanation as it may mistakenly appear that it doesn't matter whether patients visit the clinician simultaneously or one at a time. Suppose two patients, Tom and Joe, visit the clinician together, and the clinician obtains fMRI data for both, $\boldsymbol{f}_\mathrm{T}$ and $\boldsymbol{f}_\mathrm{J}$. The joint post-test probability for their health conditions $c_\mathrm{T}$ and $c_\mathrm{J}$ is different from the one obtained first calculating Joe's one, say, and then Tom's using Joe's results:

$$\mathrm{p}(c_\mathrm{T}, c_\mathrm{J} |\, \boldsymbol{f}_\mathrm{T}, \boldsymbol{f}_\mathrm{J}, D, M_l) \neq \mathrm{p}(c_\mathrm{T} |\, \boldsymbol{f}_\mathrm{T}, c_\mathrm{J}, \boldsymbol{f}_\mathrm{J}, D, M_l) \times \mathrm{p}(c_\mathrm{J} |\, \boldsymbol{f}_\mathrm{J}, D, M_l). \tag{2.30}$$

This inequality can be easily verified by applying the probability product rule to the left side, and can be understood as follows. Suppose the clinician first wants to calculate the likelihood for Joe's being healthy. If Tom is schizophrenic then his fMRI result is unimportant for Joe's likelihood, owing to our assumption of independent priors (2.16). But if Tom is healthy, then his fMRI result, *which is known to the clinician*, should lead to an updated model for Joe's likelihood. The likelihood for Joe's being healthy is therefore a mixture of these two possible likelihoods, with weights proportional to the post-test probability for Tom's health condition. Thus, Joe's likelihood is affected by Tom's fMRI result even if Joe's is calculated first and

Tom's health condition is not yet known. More generally, if several patients visit the clinician simultaneously, she should order diagnostic fMRI tests for all of them at once and calculate a joint post-test probability for them, in order to make the best-informed prediction for each.

## 2.3 Discussion

### 2.3.1 Summary

The diagnosis of a medical condition is a complex process that takes in a variety of judgements and evidence from the clinician and from any diagnostic tests available to her. Bayesian probability theory has found wider acceptance in medicine because it can consistently combine and frame such judgements and evidence (Goodman, 1999; Davidoff, 1999; Greenland, 1998). Formulae (2.1)–(2.3) show the basic scheme of how the clinicians' judgements and the results of diagnostic tests are combined (Sox et al., 2013). The role of a diagnostic test is not simply to give a dichotomic answer, e.g. healthy/ill, but a *likelihood* for each health condition, to be combined into this scheme together with the likelihoods from other tests. The final probability obtained from these likelihoods is finally used by the clinician to decide upon a course of action, e.g. dismiss/treat (Sec. 2.4.3).

The values of likelihoods from such tests need to be determined from a set of training set of data for each health condition. In this work we have discussed how to determine the likelihoods when the diagnostic test and training set consist of fMRI data, considering for concreteness the case of schizophrenia as the disease in question, and by using real fMRI data from healthy and schizophrenic patients (Sec. 2.2.1). We derived them step by step from first principles through a sequence of assumptions:

(a) **Partial exchangeability** with respect to the health conditions, explained in Sec. 2.2.2. We believe this assumption to be very natural in medical diagnostics. By itself it already leads to a specific expression for the likelihood, Eq. (2.8), although this expression is very difficult to compute.

(b) **Sufficiency** of an empirical statistics of a reduced set of fMRI data (Secs. 2.2.3–2.2.4). We believe such kind of assumption to be sound and at least approximately true when neurologically motivated, and moreover it provides a bridge between biophysical considerations and the specification of probabilities. It leads to a likelihood, Eq. (2.11), amenable to numerical or analytic computation.

(c) **Prior knowledge** of the empirical statistics for the different health conditions (Sec. 2.2.4.4). An assumption of this kind is always necessary, especially with small training data sets; but it affects our inferences less and less as our training data accumulate.

In our study, for (b) we specifically assumed the sufficiency of the first and second moments of some functions (Sec. 2.2.4) of the functional-connectivity weights obtained from the fMRI data (Sec. 2.2.3). Our focus on connectivities was neurologically motivated, but our choice of first and second moments of particular functions was made for mathematical simplicity, in

order to illustrate the method. Our choice of roughly flat, conjugate priors for assumption (c) was also made for the sake of mathematical simplicity and illustration (Sec. 2.2.4.4).

Notwithstanding our simple, mainly illustrative choices, we obtained good diagnostic performances, briefly discussed in the next section. To assess these performances and the relative predictive power of different choices in assumption (b) above, we presented several criteria, based on decision theory (Sec. 2.2.5). These criteria also help in understanding whether the training of the likelihoods has stabilized (Sec. 2.2.5). We observed that in this kind of study decision-theoretical criteria can yield results in seeming contrast with Bayesian model-comparison criteria, like weights of evidence and Bayes factors (Sec. 2.2.5.3); but this contrast is understandable and is not a sign of inconsistency.

We emphasize that the analysis and conclusions presented here are just illustrations of the general method, and are not meant to be used for real diagnoses: as explained in Secs. 2.2.4.2 and 2.2.5.4, for a real application we would first need to investigate better-informed sufficient statistics, realistic decisions available to the clinician, the latter's utilities, and realistic pre-test probabilities. In this study we used simple default values for illustrative purposes only.

### 2.3.2 Comparison with other studies and methods

In this study we used a rather naive approach, explained in Sec. 2.2.3, to select a subset of brain connections for our analysis. That approach can be criticized in two different ways. First, it ignores the fact that if distributions are narrow, the means can be close together without much overlap, and that such distributions are likely to be more informative for the purposes of discrimination. This could be improved, for example, by taking the minimal area of the distributions' overlap (dark red area in Fig. 2.1) instead. Second, we did not restrict our search for suitable connections and associated brain areas to the areas that are known to be part of resting-state networks identified in previous studies, like the default mode network. Our lack of specificity, though, was motivated by previous studies which have demonstrated that functional connectivity can be also measured in other task-related networks, induced by spontaneous activity (He et al., 2009), and that in resting state different activity patterns can appear in schizophrenic patients, owing e.g. to hallucinations during the scan (e.g. Shergill et al., 2000).

Despite our simplified and possibly unrealistic choices of connectivity weights, sufficient statistics, parameter priors, pre-test probabilities, and utility functions, we obtained a diagnostic performance of $80\%$, measured by leave-one-out cross-validation (Sec. 2.2.5), comparable to classification results based on fMRI data reported in previous studies (e.g., Venkataraman et al., 2012, 18 healthy + 18 schizophrenic patients; Çetin et al., 2016, 45 + 46 patients; Demirci et al., 2008, 36 + 34 patients ).

Our results also compare with classification results that we achieved using machine-learning methods. Cross-validation tests using support-vector machines (using $80\%$ of the data for training and $20\%$ for testing in a randomized iterative way) also yielded around $80\%$ of correct diagnoses (data not shown).

But, as explained in the Introduction of this chapter (points I., II.), methods like these, which simply classify or are deterministic, do not fit the clinician's decision-theoretical problem: they cannot be combined with other diagnostic tests and do not fit a general decision-theoretic approach – cf. Secs. 2.2.5.1, 2.4.3. Most machine-learning methods (Bishop, 2006; Murphy, 2012) are thus ruled out.

There is no real contrast, however, between machine-learning methods and the method presented here: machine-learning algorithms can be interpreted as convenient, fast approximations of Bayesian methods (see e.g. the explicative image in Huszár, 2017), often combined with default utility functions and decision rules (Murphy, 2012; MacKay, 2003, 1992a,e,d,c,b). A machine-learning classification algorithm that gives a good performance can suggest good likelihoods or parameter priors to be used in a statistical model. For example, the simplest version of a support-vector machine (Bishop, 2006, Ch. 7; Murphy, 2012, Sec. 14.5) can be interpreted as a model where the likelihood $L_{\mathrm{H}}(\boldsymbol{f}|\,\boldsymbol{\theta}_{\mathrm{H}})$ for one health condition is very large in a half of the dataspace $\boldsymbol{f} \in [-1,1]^d$ and zero in the rest, the hyperplane separating these half-spaces being determined by the parameter $\boldsymbol{\theta}_{\mathrm{H}}$. The likelihood $L_{\mathrm{S}}(\boldsymbol{f}|\,\boldsymbol{\theta}_{\mathrm{S}})$ for the other health condition is likewise large or zero in two half-spaces separated by a hyperplane determined by $\boldsymbol{\theta}_{\mathrm{S}}$; see Eqs. (2.11) in Sec. 2.2.4.1. In this model the parameter prior $\mathrm{p}(\boldsymbol{\theta}_{\mathrm{H}}, \boldsymbol{\theta}_{\mathrm{S}}|\, M, I)$ correlates the two parameters $(\boldsymbol{\theta}_{\mathrm{H}}, \boldsymbol{\theta}_{\mathrm{S}})$ in such a way that the two hyperplanes coincide and the two likelihoods have support on opposite sides. As the model is trained and the parameter prior is updated, Eq. (2.11b), the hyperplane moves in space in a way to maximize the product of the likelihoods. This corresponds to the search of an optimal separation hyperplane by the support-vector machine.

### 2.3.3 Possible improvements

Besides using more realistic pre-test probabilities, and utility values, the method presented here could be improved in several other respects, especially with regard to assumptions (b) and (c) summarized in Sec. 2.3.1.

In point (b) we assumed that particular connectivity weights are sufficient to distinguish among schizophrenic and healthy patients. These connectivities were calculated as in Secs. 2.2.3 and 2.4.1. More sophisticated choices of Regions Of Interests and functional-connectivity measures (Marrelec & Fransson, 2011; Smith et al., 2011; Wang et al., 2014; Gheiratmand et al., 2017; Demirci et al., 2008) or even integration of functional and structural imaging (Michael et al., 2010) could obviously lead to an increased predictive power. A different choice of sufficient statistics could also improve the performance. With increased computational power it would even be possible to use the full set of connectivities rather than sufficient statistics thereof – so-called "nonparametric" models (cf. Zhang et al., 2014, 2016; Nielsen et al., 2016; Kook et al., 2017).

Regarding assumption (c), in Sec. 2.2.4.3 we mentioned that parametric models may lead to probabilities that lack a closed form and are analytically intractable. The models we chose, with conjugate prior, have the advantage of having closed-form formulae, but they also restrict our choice of sufficient statistics and parameter priors. Higher predictive power

could be achieved by using other kinds of sufficient statistics, leading to likelihoods that are not generalized normals, or by using non-conjugate priors, e.g. treating means, correlations, and variances independently, as discussed by Barnard et al. (2000). In this case numerical methods are needed, such as Markov-Chain Monte Carlo sampling and integration (MacKay, 2003, Ch. IV; Murphy, 2012, Chs. 23–24). It must be kept in mind, though, that numerical methods can be computationally vastly more expensive than analytic ones. In preliminary studies that led to our present work we considered a couple of statistical models that require Monte Carlo sampling: a truncated normal and a product of beta distributions among them. The calculation of the relevant integrals for these models has vastly higher time costs than for the closed-form models presented here. For example, calculation of a posterior parameter distribution for the truncated-normal model required $17\,\mathrm{h}$ on a computer cluster; whereas the corresponding calculation for the models in the present work takes a fraction of a second on a laptop. To assess and select a model for our clinician to use, such integrals need to be computed over and over, as we explained in Sec. 2.2.5.1. The assessment of statistical models that require Monte Carlo methods can therefore lead to months of computation. Further explorations are needed in this direction.

The comparison with support-vector machines sketched at the end of the previous section shows one important assumption of our statistical models: the independence of the prior parameter distributions for the two health conditions: $\mathrm{p}(\boldsymbol{\theta}_\mathrm{H}, \boldsymbol{\theta}_\mathrm{S}|\, I) = \mathrm{p}(\boldsymbol{\theta}_\mathrm{H}|\, I) \times \mathrm{p}(\boldsymbol{\theta}_\mathrm{S}|\, I)$, see Eq. (2.16). Statistical models of which support-vector machines are approximations clearly cannot make this assumption. The performance of the model might improve by using a non-independent joint prior distribution, which allows training data for one health condition to influence the parameter distribution for the other. In the case of the generalized-normal models we examined, this can be achieved – whilst preserving their computational convenience – by taking a weighted average of several values of prior coefficients (2.33). This is known as a hierarchical model (Bernardo & Smith, 2000, Sec. 4.6.5).

The possibilities for improvement listed in these last paragraphs suggest that the method we have presented here, hinging on first principles, has great potential for applications and for development in different directions; moreover this is not to the exclusion of other methods, but assimilating their principles and advantages.

## 2.4 Methods

### 2.4.1 Data preprocessing

Preprocessing of the rfMRI images is carried out using the FMRIB Software Library tools (FSL, v5.08: Jenkinson et al., 2012; Smith et al., 2005) and consists of the following steps: removal of the first ten image volumes, leaving the remaining 130 volumes for further data processing; removing non-brain tissue (BET: Smith, 2002b); motion correction (MCFLIRT: Jenkinson et al., 2002); spatial smoothing with a $6\,\mathrm{mm}$ full width at half maximum normal kernel; temporal low-pass filtering with a cut-off frequency of $0.009\,\mathrm{Hz}$; white matter and cerebrospinal fluid regression (FSL regfilt, MELODIC).

For each subject we first linearly register the rfMRI image first to the structural, skull-removed image (image segmentation for skull removing with SPM8, Wellcome Department of Cognitive Neurology, London, UKFSL; linear registration with FSL/FLIRT: Jenkinson & Smith 2001; Jenkinson et al. 2002) and then, through a non-linear mapping, to the MNI standard brain (non-linear registration with Advanced Normalization Tools (ANTs: Avants et al., 2011); MNI 152 standard brain, non-linear 6th generation (Grabner et al., 2006). Regions of interest (ROIs) of the resulting functional image in standard space are extracted such that they match the 94 regions identified by the Oxford lateral cortical atlas with a probability above $50\%$ (Desikan et al., 2006). The temporal mean signals across the voxels in each ROI are used to calculate the functional connectivity measured based on the Pearson correlation coefficient.

### 2.4.2 The normal model with conjugate prior

This statistical model, denoted in this section by $M$, is amply discussed in the literature (Gelman et al., 2014, Sec. 3.6; Minka, 2001; Murphy, 2007); here we only give a summary.

Its likelihood is a normal distribution

$$L(\boldsymbol{f}|\,\boldsymbol{\lambda},\boldsymbol{\Lambda},M,I) = \mathrm{N}[l(\boldsymbol{f})|\,\boldsymbol{\lambda},\boldsymbol{\Lambda}]\,l'(\boldsymbol{f}) \tag{2.31}$$

with mean $\boldsymbol{\lambda}$ and covariance matrix $\boldsymbol{\Lambda}$.

The prior distribution for the parameters $(\boldsymbol{\lambda},\boldsymbol{\Lambda})$ is a normal-inverse-Wishart distribution, i.e. the product of a normal distribution for $\boldsymbol{\lambda}$ and an inverse-Wishart matrix distribution (Gupta & Nagar, 2000, Sec. 3.4; Tiao & Zellner, 1964; Bernardo & Smith, 2000, Sec. 3.2.5) for $\boldsymbol{\Lambda}$:

$$\mathrm{p}(\boldsymbol{\lambda},\boldsymbol{\Lambda}|\,\kappa_0,\boldsymbol{\delta}_0,\nu_0,\boldsymbol{\Delta}_0,M_l,I) = \mathrm{p}(\boldsymbol{\lambda}|\,\boldsymbol{\Lambda},\kappa_0,\boldsymbol{\delta}_0,M_l,I)\times\mathrm{p}(\boldsymbol{\Lambda}|\,\nu_0,\boldsymbol{\Delta}_0,M_l,I), \tag{2.32}$$

with

$$\begin{aligned}\mathrm{p}(\boldsymbol{\lambda}|\,\boldsymbol{\Lambda},\kappa_0,\boldsymbol{\delta}_0,M_l,I) &= \mathrm{N}(\boldsymbol{\lambda}|\,\boldsymbol{\delta}_0,\boldsymbol{\Lambda}/\kappa_0),\\ \mathrm{p}(\boldsymbol{\Lambda}|\,\nu_0,\boldsymbol{\Delta}_0,M_l,I) &= \mathrm{Wishart}^{-1}(\boldsymbol{\Lambda}|\,\nu_0,\boldsymbol{\Delta}_0)\propto\det(\boldsymbol{\Lambda})^{-\frac{\nu_0+d+1}{2}}\,\exp\bigl(-\tfrac{1}{2}\,\mathrm{tr}\,\boldsymbol{\Delta}_0\boldsymbol{\Lambda}^{-1}\bigr).\end{aligned} \tag{2.33}$$

It should be noted how $\boldsymbol{\Lambda}$ appears as parameter in the distribution for $\boldsymbol{\lambda}$, so their distributions are not independent. The composite distribution depends on two scalar, one vector, and one matrix coefficients $(\kappa_0,\boldsymbol{\delta}_0,\nu_0,\boldsymbol{\Delta}_0)$.

This prior parameter distribution retains the same form when it is conditioned on the data $(\boldsymbol{f}_i)$ of $n$ patients, becoming a posterior parameter distribution with updated coefficients $(\kappa,\boldsymbol{\delta},\nu,\boldsymbol{\Delta})$ depending on the prior ones and on the sufficient statistics:

$$\begin{aligned}&\kappa=\kappa_0+n, && \nu=\nu_0+n,\\ &\boldsymbol{\delta}=\frac{\kappa_0\,\boldsymbol{\delta}_0+n\,\overline{\boldsymbol{f}}}{\kappa_0+n}, && \boldsymbol{\Delta}=\boldsymbol{\Delta}_0+n\;\mathrm{Cov}(\boldsymbol{f})+\frac{\kappa_0\,n}{\kappa_0+n}\bigl(\overline{\boldsymbol{f}}-\boldsymbol{\delta}_0\bigr)\bigl(\overline{\boldsymbol{f}}-\boldsymbol{\delta}_0\bigr)^{\intercal}.\end{aligned} \tag{2.34}$$

The main features of the normal-inverse-Wishart distribution for $(\boldsymbol{\lambda}, \boldsymbol{\Lambda})$ are these:

$$\begin{aligned} \boldsymbol{\lambda}\text{:} &\quad \text{mean \& mode} = \boldsymbol{\delta}, \qquad \text{covariances} = \frac{1}{\kappa\,(\nu - d - 1)}\boldsymbol{\Delta}; \\ \boldsymbol{\Lambda}\text{:} &\quad \begin{cases} \text{mean} = \dfrac{1}{\nu - d - 1}\boldsymbol{\Delta}, \qquad \text{mode} = \dfrac{1}{\nu + d + 2}\boldsymbol{\Delta}, \\ \text{diagonal variances} = \dfrac{2}{(\nu - d - 3)(\nu - d - 1)^2}(\boldsymbol{\Delta})_{kk}{}^2. \end{cases} \end{aligned} \tag{2.35}$$

These formulae above say that the uncertainty in the location parameter $\boldsymbol{\lambda}$ decreases as $\kappa$ and $\nu$ increase for fixed $\boldsymbol{\Delta}$, and the uncertainty in the matrix scale parameter $\boldsymbol{\Lambda}$ decreases with increasing $\nu$. When $\nu = d + 1$ the marginal distributions for the correlations of $\boldsymbol{\lambda}$ are uniform (Gelman et al., 2014, Sec. 3.6; Barnard et al., 2000, Sec. 2.2). Because of these properties, a "vaguely informative" parameter distribution should have small $\kappa_0$ and $\nu_0$ (Minka, 2001; Murphy, 2007).

When the likelihood (2.31) and the parameter prior (2.32), updated with (2.34), are multiplied and the parameters are integrated, the resulting distribution for $\boldsymbol{f}$ is a multivariate t distribution (Kotz & Nadarajah, 2004; Minka, 2001; Murphy, 2007)

$$\begin{aligned} \mathrm{p}[\boldsymbol{f}|\,(\boldsymbol{f}_i), \kappa_0, \boldsymbol{\delta}_0, \nu_0, \boldsymbol{\Delta}_0, M_l, I] &\equiv \mathrm{p}(\boldsymbol{f}|\,\kappa, \boldsymbol{\delta}, \nu, \boldsymbol{\Delta}, M_l, I) \\ &= \mathrm{t}\Big[l(\boldsymbol{f})\big|\ \nu_{\mathrm{H}} - d + 1, \boldsymbol{\delta}_{\mathrm{H}}, \tfrac{\kappa_{\mathrm{H}}+1}{\kappa_{\mathrm{H}}\,(\nu_{\mathrm{H}}-d+1)}\boldsymbol{\Delta}_{\mathrm{H}}\Big] \prod_k l'(f_k) \end{aligned} \tag{2.36}$$

with $\nu - d + 1$ degrees of freedom, mean $\boldsymbol{\delta}$, scale matrix $\frac{\kappa+1}{\kappa\,(\nu-d+1)}\boldsymbol{\Delta}$, and covariance matrix $\frac{\kappa+1}{\kappa\,(\nu-d-1)}\boldsymbol{\Delta}$.

### 2.4.3 Decision theory and utility

Once we have the post-test probabilities $(p_{\mathrm{H}}, p_{\mathrm{S}})$ for the possible health conditions of a patient given the fMRI data, there remains to decide upon a course of action. This is the domain of decision theory (Raiffa & Schlaifer, 2000; Jaynes, 2003, Chs. 13, 14; Sox et al., 2013, Chs. 6, 7).

Suppose we have only two courses of action: treat T or dismiss D. A decision-theoretical analysis needs, besides the probabilities for the health conditions, also the utilities (or costs) of choosing an action given the patient's true health condition. For example, treatment of a healthy patient could harm the latter, or it could be innocuous. With two courses of action and two health conditions we have four utilities $u_{\text{decision|condition}}$:

| | healthy | schizophrenic | |
|---|---|---|---|
| dismiss | $u_{\mathrm{D|H}}$ | $u_{\mathrm{D|S}}$ | $u_{\mathrm{D|H}} > u_{\mathrm{T|H}}$, |
| treat | $u_{\mathrm{T|H}}$ | $u_{\mathrm{T|S}}$ | $u_{\mathrm{T|S}} > u_{\mathrm{D|S}}$. |

(2.37)

Typically $u_{\mathrm{D|H}} > u_{\mathrm{T|H}}$ and $u_{\mathrm{T|S}} > u_{\mathrm{D|S}}$, and $u_{\mathrm{D|H}}$, $u_{\mathrm{T|S}}$ are positive and $u_{\mathrm{T|H}}$, $u_{\mathrm{D|S}}$ negative if we appropriately shift the zero of our measurement units.

The expected utilities for dismissal and treatment are therefore

$$\mathrm{E}(u_{\mathrm{D}}) = u_{\mathrm{D|H}}\, p_{\mathrm{H}} + u_{\mathrm{D|S}}\, p_{\mathrm{S}}, \qquad \mathrm{E}(u_{\mathrm{T}}) = u_{\mathrm{T|H}}\, p_{\mathrm{H}} + u_{\mathrm{T|S}}\, p_{\mathrm{S}}. \tag{2.38}$$

Decision theory says the clinician ought to chose the action having maximum expected utility. For example, she dismisses the patient if $\mathrm{E}(u_{\mathrm{D}}) > \mathrm{E}(u_{\mathrm{T}})$, that is if

$$p_{\mathrm{S}} < \frac{u_{\mathrm{D|H}} - u_{\mathrm{T|H}}}{u_{\mathrm{D|H}} - u_{\mathrm{T|H}} + u_{\mathrm{T|S}} - u_{\mathrm{D|S}}}. \tag{2.39}$$

# Chapter 3

# On the extraction and analysis of graphs from resting-state fMRI to support a correct and robust diagnostic tool for Alzheimer's disease

The diagnosis of Alzheimer's disease (AD), especially in the early stage, is still not very reliable and the development of new diagnosis tools is desirable. A diagnosis based on fMRI is a suitable candidate, since fMRI is non-invasive, readily available, and indirectly measures synaptic dysfunction, which can be observed even at the earliest stages of AD. However, previous attempts to analyze graph properties of resting state fMRI data are contradictory, presumably caused by methodological differences in graph construction. This comprises two steps: clustering the voxels of the functional image to define the nodes of the graph, and calculating the graph's edge weights based on a functional connectivity measure of the average cluster activities. A variety of methods are available for each step, but the robustness of results to method choice, and the suitability of the methods to support a diagnostic tool, are largely unknown. To address this issue, we employ a range of commonly and rarely used clustering and edge definition methods and analyze their graph theoretic measures (graph weight, shortest path length, clustering coefficient, and weighted degree distribution and modularity) on a small data set of 26 healthy controls, 16 mild cognitive impairment and 14 Alzheimer's disease. We examine the results with respect to statistical significance of the mean difference in graph properties, the sensitivity of the results to model and parameter choices, and relative diagnostic power based on both a statistical model and support vector machines. We find that different combinations of graph construction techniques yield contradicting, but statistically significant, relations of graph properties between health conditions, explaining the discrepancy across previous studies, but casting doubt on such analyses as a method to gain insight into disease effects. The production of significant differences in mean graph properties turns out not to be a good predictor of future diagnostic capacity. Highest predictive power, expressed by largest negative surprise values, are achieved for both atlas-driven and data-driven clustering (Ward clustering), as

long as graphs are small and clusters large, in combination with edge definitions based on correlations and mutual information transfer.

## 3.1 Introduction

The two major challenges in Alzheimer disease (AD) research consist in firstly, finding an effective treatment that at least slows down the disease progress, and secondly, developing diagnostic tools that can not only detect the disease at the earliest stage, during which no symptoms related to cognitive deficits are apparent (Sperling et al., 2011), but also provide information into the progression of the disease. For the latter challenge it is particular desirable that the tools can be deployed within the existing medical infrastructure (i.e. not requiring specialized machinery or lab procedures), such that it is feasible to scan a wide range of the elderly population. Diagnosis procedures currently in use include psychological tests, detection of abnormal concentrations of disease specific biomarkers (Amyloid-$\beta$ , Tau proteins) in cerebrospinal fluid and analysis of structural magnetic resonance images (MRI).

Although abnormalities of Amyloid-$\beta$ concentrations are proposed to be the earliest disease indicator, they are not very reliable in disease prognosis. Moreover, the changes in Amyloid-$\beta$ concentrations show the strongest increase in the preclinical phase, and are thus uninformative with respect to the further progression of the disease. Tau pathology, which probably spreads along functional networks (Hoenig et al., 2018) better predicts cognitive deficits and progression of the disease (Nelson et al., 2012). However, the two methods measuring Amyloid-$\beta$ and Tau concentrations, lumbar puncture and PET are invasive (Sperling et al., 2011; Schroeter et al., 2009).

Possibly, synaptic dysfunction, another disease marker, corresponds to the onset of AD even before Amyloid-$\beta$ pathology starts. Additionally, as it gradually worsens throughout the course of the disease, it could serve as diagnostic marker for all stages of AD. Dysfunction of synapses can be indirectly measured via invasive FDG-PET and non-invasive functional MRI, which might directly be combined with structural MRI scans (Sperling et al., 2011; Schroeter et al., 2009). However, a diagnostic framework based on functional MRI has yet to be established.

Although many fMRI studies have investigated changes of functional activity in AD (for a review see Dennis & Thompson, 2014), there is no consensus about which information should be used. Such studies typically examine disrupted cortical connectivity, either locally, considering single brain areas (e.g. Dillen et al., 2017) and their embedding in the network, or globally, analyzing the entire constructed brain graph and the statistics of its graph properties (Gits, 2016).

We argue that in order to develop a robust diagnosis tool applicable to all disease stages, it is preferable to consider global graph properties for the following reasons. First, global graph properties seem to be more robust across sessions; consequently, changes in these properties over time are more likely to reflect disease progression than statistical fluctuations (Wang et al., 2014; Telesford et al., 2010). Second, not all disease progression follows a stereotypical

pattern. Whereas structural evidence of AD is typically found predominantly in entorhinal cortex and hippocampus, in atypical cases atrophy occurs primarily in other areas, such as posterior cortex (Johnson et al., 2012). These atypical cases might be better captured by global properties, since they make use of the entire information provided by the brain. Furthermore, analyzing the statistics of graph properties rather than comparing the properties of single nodes allows the use of data-driven brain clustering, which results in different numbers and locations of brain clusters for each individual.

However, it is challenging to investigate the informativeness of global graph properties due to the innumerable methods of graph construction methods, comprising both the clustering of the voxels to define the graph's nodes, and the definition of functional connectivity to define its edges. Across the range of previous studies investigating graph properties in AD, a wide variety of methodological approaches for graph construction and properties assessment have been applied and are probably a major source of contradictory observations, such as the comparative length of the shortest path in AD subjects with respect to control being reported in two recent studies as both shorter (Zhao et al., 2012) and longer (Sanz-Arigita et al., 2010).

It is a further challenge to identify an appropriate evaluation method that not only enables us to compare the different graph construction methods, but also permits the results to be combined with other information indicating the probability of a particular health condition. This means that pure classifiers, although they achieve high discrimination performance (Khazaee et al., 2015, 2017) do not meet these requirements because they return a group membership ('AD', 'MCI' or 'control') and not a probability that can be combined with the results of other diagnostic tests (e.g. derived from Amyloid-$\beta$ concentration measures) or individual patient risk factors (Porta Mana et al., 2018).

In this article, we address these issues by presenting a methodology for determining which combination of techniques to extract and analyze graphs from resting state functional magnetic resonance imaging (fMRI) data provides the best basis for a diagnosis tool, assuming a given initial data set. Here, we apply our methodology to a small data set consistig of 26 control (C) elderly patients without any indication of any form of dementia or other cognitive problems, 16 mild cognitive impaired (MCI) subjects and 14 patients suffering from Alzheimer disease (AD) (Dillen et al., 2017). We evaluate the combinations of graph construction and analysis methods using a statistical model that partly compensates for the small data set and also yields probabilities rather than classifications, thus permitting the results to be combined with other probabilities, as discussed above. In addition, we evaluate the graph construction techniques with respect to robustness of results to method configuration parameters and similarity of results across different techniques.

Note that our aim here is not to demonstrate superior classification (for which our data set is in any case too small) or to propose a particular combination of techniques as optimal (as this may vary between settings), but primarily to provide a principled way for determining an appropriate combination of techniques for a given data set, and secondarily to highlight the sensitivity of graph theoretical analysis to the details of graph construction.

To understand how different methods for constructing graphs affect the resultant graph properties, and thus the ability to distinguish between patient groups, we evaluate a range

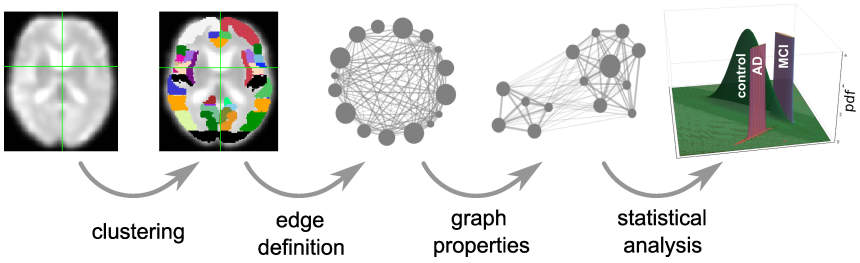


**Figure 3.1: Overview of intermediate steps for graph construction, properties derivations and statistical analysis**. Each picture illustrates the result of a processing step starting from the preprocessed function image (very left), which is clustered into regions, used as the nodes of the graph (second image). The averaged fMRI activity of each region is then used to calculate the edges of the graph (third image) and based on the calculated graph properties (fourth image) of all graphs, the statical analysis estimates the probability density functions (pdf) of the three health conditions (last image) that are necessary for the evaluation of diagnostic performance based on the negative surprise measure. For the first three steps of the pipeline we investigate a range of different methods, see Secs. 3.2.1, 3.4.3, 3.4.4 and 3.4.5 for details.

of standard and non-standard methods to construct the graphs. The first step in graph construction consists in clustering adjacent voxels, such that the activity of the resulting region can be expressed by the average of time varying signal of the selected voxels (see Fig. 3.1). The decision as to which voxels form a cluster is often based on atlases established for a standard brain with predefined brain regions. In order to map this standard atlas to the functional image or vice versa, registration algorithms are used. Problematic in this step, especially for subjects potentially suffering from neurodegenerative diseases, is the inhomogeneous shrinkage of the brain, which hampers a correct registration (Liu et al., 2017). In addition, individual brain regions derived from standard brain templates are likely to execute several cognitive processes in parallel, such that averaging the activity across the voxels of these functional inhomogeneous regions is not justified (Marrelec & Fransson, 2011). We therefore also include activity driven algorithms, namely region growing and selection (RGS, Lu et al., (2003)) and Ward clustering, into our evaluation.

In the second step in graph construction, functional connectivity values are calculated based on the averaged signal of the regions. In most studies this is carried out based on the Pearson correlation coefficient, restricting the functional connectivity to non-directional connections. Here we cover a broader range of possible measures in the time domain: linear, non-linear model-free and model-based (Wang et al., 2014) that depending on their exact realization result in directed or undirected graphs.

We then calculate a variety of graph measures on the single nodes (weighted degree, cluster coefficient, closeness centrality), edges (weights, shortest path) and the entire graph (modularity). As several of these measures are only well-defined for binary graphs, many studies binarize the weighted graphs obtained from the previous steps into binary graphs, by setting weights above an arbitrary threshold $w_{\min}$ to 1, and those below it to 0 (e.g. Supekar et al., (2008)). The drawback here is that there is no validation for an optimal threshold, and information that might be relevant in AD may be lost. To investigate this problem, we analyze

the dependence of graph theoretic measures on $w_{\min}$, setting the weights below it to 0 but leaving the values above unchanged.

To assess the suitability of combinations of graph construction and analysis methods to inform a diagnosis tool, we set up a statistical analysis based on a training data set of known health conditions (healthy controls, mild cognitive impairment, and Alzheimer disease), see Sec. 3.4.6. The diagnostic usefulness of the analysis pipeline is then defined as the performance of the model against a labeled test data set. A model with good performance can ultimately be employed in a clinical setting, to assess the probability that a patient has one of the three health conditions. For a more complete discussion of the development and use of the statistical model, see (Porta Mana et al., 2018).

In this study we use a statistical model constructed from the following working hypothesis: the empirical means and correlations of graph data from previous patients with a given health condition are sufficient to predict the graph data of a new patient with that same health condition. This is a partially exchangeable model by sufficiency, and the resulting likelihood is a multivariate t distribution (Porta Mana et al., 2018), described in Sec. 3.4.6. To assess which graph constructions have the greatest predictive power, we calculate their log-probabilities or *negative surprises* (Bartlett, 1952; Good, 1956, 1957a). To validate this approach, we also compare the results of the negative suprise with the classification performance achieved by a support-vector machine (SVM).

Our results show that clustering resulting in small graphs with large clusters (Ward and atlas-based clustering) achieve highest negative surprises (and best SVM classification performance). Similarly, amongst the edge definition techniques, model-free methods (linear and non-linear correlations, mutually information transfer) obtain the highest negative surprise values. Conversely, calculating the graph's edge weights according to transfer entropy (model based) achieves limited diagnostic power but the ordering of the individuals based on their average graph properties is very robust towards the applied clustering method and choice of algorithm specific parameters. We further demonstrate that significant differences in the means of graph properties are very sensitive to method choice and to parameterization choices for a given method. Therefore such results, if taken at face value and not validated by alternate methods, may well be artefactual and not provide insight into the effects of a disease. Interestingly, the presence of significant differences in mean values of graph properties is not a reliable predictor of later diagnostic performance. In particular, atlas clustering results in only few significant differences but reaches the highest values for negative surprises and the best classification scores for the SVM. Finally, we show that the effect of setting a threshold on the graphs edge weights has only marginal effect on the negative surprise as long as threshold values are small.

## 3.2 Results

### 3.2.1 Graph construction

#### 3.2.1.1 Vertex definition by means of clustering

A universal property of the clustering algorithms examined here is the existence of a control parameter that regulates how the clusters are formed, and thus preserves a certain feature (or features) of the clusters. In atlas-based clustering, the preserved features are the number of clusters and the number of voxels per cluster. In Ward clustering, the number of resulting clusters is fixed, which we violate to a small extent by deleting very small clusters. In RGS, the homogeneity of each cluster is preserved. The freedom that each of the algorithms leaves to the non-regulated features can either be considered as drawback of the algorithm, because it makes graphs less easily comparable, or as an additional feature that might even improve the diagnosis performance.

Fig.3.2 shows the number of nodes/clusters, the average number of voxels per node and the average heterogeneity of the nodes for two configurations of the RGS algorithm, four configurations of the Ward algorithm, and the atlas algorithm (see Sec. 3.4.3 and Table 3.3). Most strikingly, the node properties vary far more with respect to the clustering method chosen than with respect to the health condition.

By construction, the number of nodes for atlas clustering are the same for all individuals, and are the smallest over all the clustering methods (top panel). In Ward clustering the number of clusters is an algorithmic specific parameter; it is not constant in Fig.3.2 because we additionally include a parameter enforcing a minimum cluster size. Thus, the number of nodes for Ward clustering decrease as the minimum number of voxels per cluster $p$ increases from $10$ for 'ward1' to $25$ in 'ward4'. In RGS clustering we do not have such algorithmic specific restrictions and the number of clusters is defined by the voxel dynamics. A consequence of this is that the number of clusters per graph are more widely spread.

The average number of voxels per cluster, shown in the middle panel of Fig.3.2, is unsurprisingly negatively correlated with the number of clusters. For purposes of comparison, the number of voxels for atlas clustering was first calculated for the standard space and then downscaled in proportion to the relation of the total number of voxels present in functional space to those in standard space. An inverse correlation can also be seen in the width of the distributions between the top two panels, for the non-atlas methods. In the case of the RGS clustering, this can be explained by the fixation of the heterogeneity to one (see bottom panel of Fig.3.2), leading to quite homogeneous numbers of voxels per cluster, but to a wide range of the number of nodes, namely from $200$ to $1200$. Since this range is so large, it could be argued that graph properties that depend on this number would not be comparable in a meaningful fashion. In order to take care of such dependencies, we include the number of nodes in our statistical analysis (Sec. 3.4.5). For Ward clustering we can observe that the numbers of nodes is inversely correlated not only with the average number of nodes its variability, but also with the average heterogeneity and its variability. We observe the highest

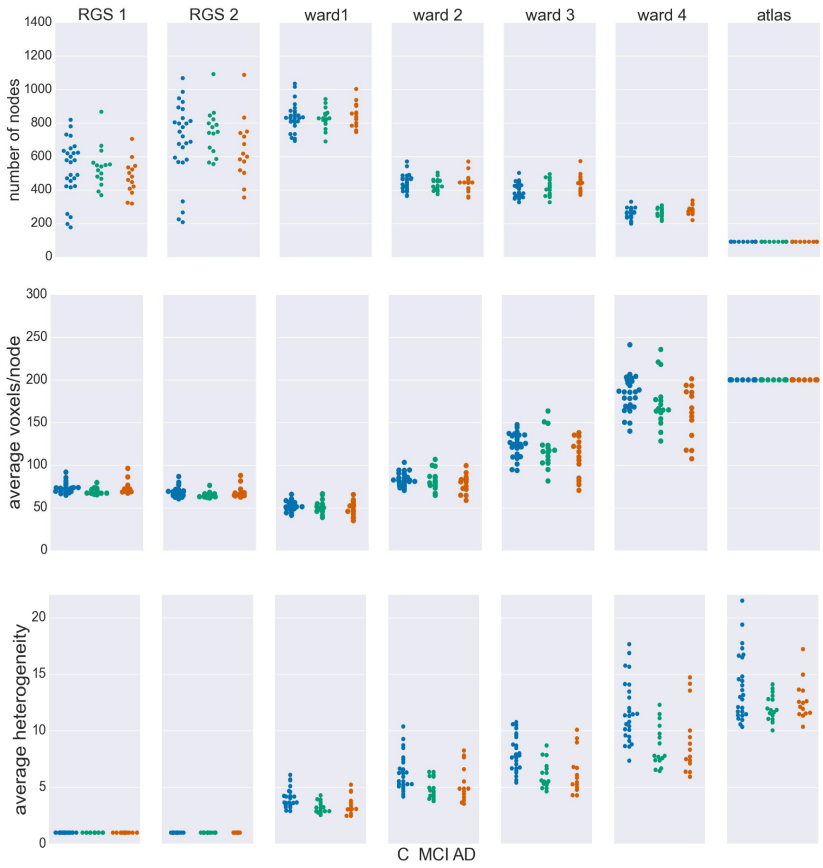


**Figure 3.2: Node properties across different clustering algorithms.** For each of the seven clustering methods detailed in Sec. 3.4.3 and Table 3.3, and each subject categorized in the health conditions: control (C, blue dots), mild cognitive impairment (MCI, green dots), and Alzheimer's disease (AD, orange dots) we calculate the total number of nodes/clusters generated (upper panel), the average number of voxels per node (middle panel), and the average cluster heterogeneity.

degree of heterogeneity for atlas clustering, presumably due to the high number of voxels per cluster.

Comparing node properties between the classes of clustering methods, atlas and ward4 clustering seem to be quite similar, which suggests they might result in similar graph properties and diagnosis performance. In particular, we note that these methods reveal a much smaller heterogeneity for the MCI group than for the control and AD groups.

#### 3.2.1.2 Edge definition by means of functional connectivity

The edges of the graphs are constructed in four different ways, described in detail in Sec 3.4.4. Linear correlations ($corr$) are based on the Pearson correlation coefficient; non-linear correlations ($H_2$) result from a non-linear fit of piecewise linear correlations; mutual information transfer ($MIT$) measures the amount of shared information between two time varying signals and transfer entropy ($TE$) describes in how far the future uncertainty is reduced by the preceding activity of the considered pair of nodes. As with the clustering algorithms described in the previous section, we defined differently parameterized variants of these four classes of technique (e.g. generating directed $D$ or undirected $U$graphs) which are listed in Table 3.4.

| combination | $\hat{w}_C$ | $\hat{w}_{MCI}$ | $\hat{w}_{AD}$ |
|---|---|---|---|
| ward $corr$ | $0.328 \pm 0.021$ | $0.337 \pm 0.04$ | $0.315 \pm 0.023$ |
| RGS $corr$ | $0.405 \pm 0.076$ | $0.363 \pm 0.049$ | $0.397 \pm 0.113$ |
| atlas $corr$ | $0.319 \pm 0.02$ | $0.334 \pm 0.054$ | $0.307 \pm 0.022$ |
| ward $H_2$ | $0.443 \pm 0.18$ | $0.398 \pm 0.081$ | $0.414 \pm 0.057$ |
| RGS $H_2$ | $0.452 \pm 0.1$ | $0.471 \pm 0.126$ | $0.493 \pm 0.179$ |
| atlas $H_2$ | $0.36 \pm 0.057$ | $0.352 \pm 0.039$ | $0.355 \pm 0.042$ |
| ward $MIT$ | $0.201 \pm 0.004$ | $0.2 \pm 0.007$ | $0.197 \pm 0.004$ |
| RGS $MIT$ | $0.221 \pm 0.026$ | $0.204 \pm 0.011$ | $0.218 \pm 0.037$ |
| atlas $MIT$ | $0.196 \pm 0.003$ | $0.197 \pm 0.008$ | $0.193 \pm 0.003$ |
| ward $TE$ | $0.163 \pm 0.013$ | $0.158 \pm 0.015$ | $0.156 \pm 0.018$ |
| RGS $TE$ | $0.158 \pm 0.026$ | $0.149 \pm 0.02$ | $0.152 \pm 0.042$ |
| atlas $TE$ | $0.163 \pm 0.016$ | $0.17 \pm 0.011$ | $0.165 \pm 0.011$ |

**Table 3.1: Mean and standard deviation of edge weight across different edge definitions.** Means and standard deviations (in brackets) are taking across the average edge weight of every individual graph in a health condition. Highest mean edge weights for each combination across the three health conditions are highlighted in gray.

For each combination of vertex (RGS, Ward or atlas) and edge definition technique ($corr$, $H_2$, $MIT$, $TE$), we averaged over the weights generated in each health condition for each variant of both techniques. For example, for the combination of region growing and transfer entropy (RGS $TE$) we averaged over all combinations of clustering implementation (RGS1 and RGS2) and edge detection ($BTEU1$, $BTEU2$, $BTED1$, $BTED2$ ). The results are shown in Table 3.1 and exhibit a high variability in the mean connection weights. For instance, the combination RGS $TE$ yields a mean weight of $0.152$ for controls, which is three times lower than the maximum mean weight of $0.452$ obtained by the RGS $H_2$ combination. In particular, RGS clustering yields higher values compared with Ward and atlas clustering for

model-free edge definitions ($corr$, $H_2$, $MIT$). The smallest values are obtained for $TE$ . As a consequence, even small thresholds e.g. $w_{\min} = 0.3$ already cause $TE$ graphs to disintegrate. Accordingly, not all graph properties can be calculated and used for statistical analysis, as shown in Sec. 3.2.3.

It is also notable that there is no systematic relationship between the three health conditions - for RGS $corr$, the control graphs have the highest mean weight, for RGS $H_2$ the AD graphs have the highest weight, and for atlas $corr$ it is the MCI graphs. These results demonstrate that conclusions drawn on health conditions based on weight statistics should be treated with suspicion, as the outcome can be strongly influenced by the method of calculation. A

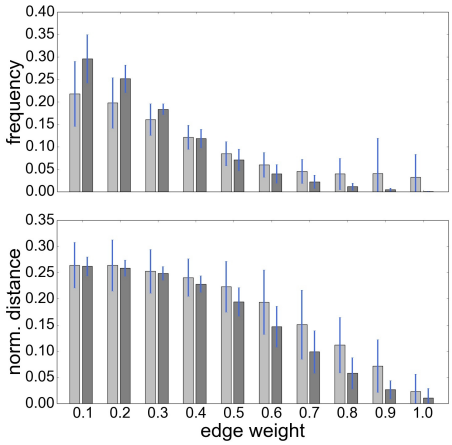


**Figure 3.3: RGS clustering yields stronger long-range connection then Ward clustering.** Frequency (upper panel) and connection distance normalized to maximum graph distance (lower panel) across a range of graph edge weights calculated based on $BcorrU1$ for RGS1 (light gray bars) and ward2 (dark gray) clustering . Mean values and standard deviation (blue vertical lines) are calculated across single histogram values of all subjects independent of health condition.

possible explanation for the higher weights generated by RGS clustering is that it produces a greater number of shorter distances compared with the other clustering techniques. However, although Fig.3.3 does indeed confirm that edge weights become smaller with cluster distance, it does not reveal a bias to shorter weights for RGS. In fact, the converse is true: RGS clustering yields stronger long-range connections for similar graph sizes (RGS: $379.69 \pm 147.99$, Ward: $311.43 \pm 33.59$). Therefore we conclude that connecting homogeneous clusters allows stronger long-range connections to be extracted. However, the statistics of the RGS connections has a much larger variance then the ones derived from Ward clustering. This is only partly due to the variance in the number of nodes, since even if we choose three healthy subject with similar graph size (RGS: $297 \pm 2.16$ , Ward: $297.33 \pm 6.6$), we still get a higher standard derivation for RGS clustering in the weight distribution ($\sigma_{\mathrm{RGS}}/\sigma_{\mathrm{ward}} = 1.6$).

In the following we will treat the distribution of edge weights as a graph property since it contains information about graph structure.

### 3.2.2 Graph properties

A recent survey by Gits (2016) of studies investigating graph properties in AD reveals no clear and systematic differences between heath conditions. For example, the mean clustering coefficient was found to be both significantly smaller (Supekar et al., 2008) and larger (Zhao et al., 2012) in AD compared with the aged-matched control group. We consider it likely that differences in methodology can account for many of the contradictions. However the stage of AD reached by the examined subject group may also play an important role. To investigate this aspect more closely, we examine the finding by Kim et al. (2015) that local efficiency, which corresponds to our definition of closeness centrality divided by the number of nodes in the network minus one, is increased for MCI, decreased for initial stages of AD and increased for severe AD stages with respect to the control group. The results of applying similar methods (altas-based clustering combined with $BMITU$) are shown in Fig.3.4. The top panel shows the relationship between the health conditions when closeness centrality is calculated on the full, non-thresholded graph, which reproduces the findings of Kim et al. (2015) at least for initial stages of AD. However, if the measure is calculated on the graphs' rich-clubs, i.e. the sub-graphs consisting of the nodes in the top $10\%$ for degree, a different picture emerges, as shown in the middle panel of Fig.3.4. Here, AD has an increased closeness centrality with respect to both the control and mild cognitive impairment groups, which is in line with advanced AD stages in Kim et al. (2015) .

More evidence that the outcome of a graph theoretical analysis can be highly sensitive towards the exact methodological implementation is given by considering the difference between the mean weights in the control and the AD conditions, and its significance (Sec. 3.4.7.1), in dependence on the thresholding weight used to convert weighted graphs into simple graphs. This is illustrated in the bottom panel of Fig.3.4. Here, depending on where we set the threshold for considering an edge to be relevant, results having a significance level of $p < 0.05$ can be observed for both $\hat{w}_{\mathrm{C}} > \hat{w}_{\mathrm{AD}}$ ($w_{\min} \in \{0.0,\, 0.1\}$ ) and $\hat{w}_{\mathrm{C}} < \hat{w}_{\mathrm{AD}}$ ($w_{\min} \in \{0.3,\, 0.4\}$).

Extending this analysis, we find that contradictory significant results can be obtained for a variety of graph metrics across (and sometimes within) clustering methods. Fig.3.5 shows the percentage of significant results obtained for health condition relationships in average edge weight, weighted degree, shortest path and clustering coefficient. Most strikingly, for most examined relationships, if significant differences are found at all, they are found in both directions, e.g. both for $\hat{d}_{\mathrm{C}} > \hat{d}_{\mathrm{MCI}}$ and for $\hat{d}_{\mathrm{C}} < \hat{d}_{\mathrm{MCI}}$ (weighted degree). Often a clustering algorithm favors a particular comparison direction, e.g. for the clustering coefficient, RGS clustering yields $\hat{clc}_{MCI} > \hat{clc}_{AD}$ whereas ward and atlas clustering yield$\hat{clc}_{MCI} < \hat{clc}_{AD}$ for ward and atlas-based clustering. However, we also find cases significant differences are found in both directions with approximately equal frequency, such as $\hat{sp}_C > \hat{sp}_{AD}$ and $\hat{sp}_C < \hat{sp}_{AD}$ for ward clustering. In addition, we find some clustering algorithms show a systematic behavior across metrics, e.g. for RGS $\hat{x}_C > \hat{x}_{MCI}$ with $x \in \{w,\, d,\, sp,\, clc\}$.

The largest number of significant differences are found for the comparison of controls with MCI, followed by the comparison of controls with AD. Only few significant differences of the means are found for AD and MCI. This relation among the groups is in line with the observed differences in heterogeneity observed for ward and atlas clustering, for which MCI

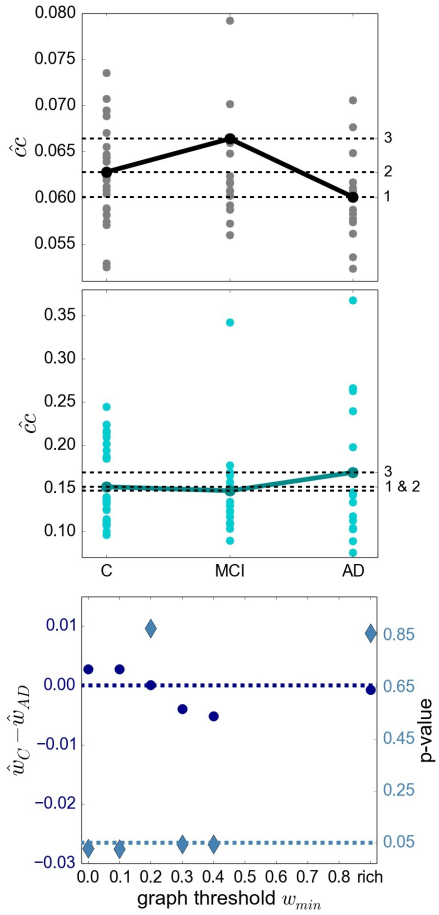


**Figure 3.4: Relationship of sub-graph properties across heath conditions is dependent on graph size.** Top panel: Average closeness centrality $\hat{cc}$ across graph nodes for whole graphs constructed with atlas $BMITU$ for the different health conditions C (right), MCI (middle) and AD (left). Each dot corresponds to the graph of an individual. Middle panel: as in top panel, but on the basis of the rich-club graphs. Bottom panel: Difference of the averaged ward1 $BMTID2$ graph weights of the control group $\hat{w}_C$ and the AD group $\hat{w}_{AD}$ (left vertical axis, blue discs) and significance of this difference (right vertical axis, turquoise diamonds) as functions of the graph thresholding value $w_{\min}$. Note: all $\hat{w}$ are positive and are only calculated as long as graphs are connected (which is the case for $w_{\min}$<0.5). Average is taken across the weights of individual graphs. The dashed dark blue line indicates $\hat{w}_{\mathrm{C}} - \hat{w}_{\mathrm{AD}}$; the dashed turquoise line indicates a significance level of $0.05$.

showed much lower heterogeneity and AD slightly lower values compared with controls (bottom panel of Fig.3.2).

Focusing on the clustering methods that bring about the most significant comparing the entire graph properties distributions results we find the highest fraction of significant differences for

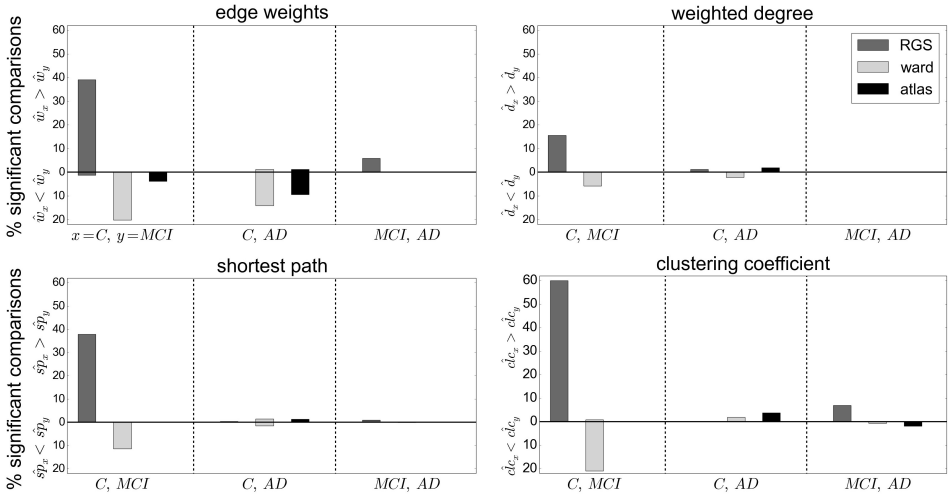


**Figure 3.5: Significant relationships in graph metrics between heath conditions dependent on clustering methods.** Percentage of significant differences for each clustering method RGS (dark gray), ward (light gray), and atlas (black) for different averaged graph properties: edge weights (top left), weighted degree (top right), shortest path (bottom left) and clustering coefficient (bottom right). Fraction of significant differences are calculated for each health condition over all graphs constructed with the corresponding clustering including all variants in parameters, edge definition techniques, thresholds and rich-club sub-graphs. The abscissa labels show which pairs of health conditions are compared (C-MCI, C-AD, MCI-AD) and the ordinate labels the according directions ('<', '>'). Significance is calculated as in the lower panel of Fig.3.4.

RGS, followed by Ward clustering. Atlas-based clustering yields only a few significant results. Fig.3.6 shows the breakdown of the proportion of significant results for each clustering method on the edge definition technique (shown in collated form in Fig.3.4). Notably, transfer entropy ($TE$) only rarely produces significant differences. All other edge definition methods show a similar fraction of significant comparisons. The highest number of significant comparisons across the different graph properties is generated by RGS clustering combined with $MIT$ .

To what extent a greater proportion of significant relationships is likely to make this graph construction method a good basis for a diagnostic tool depends on two aspects. First, the significance test is performed only on mean values, but ideally the overall distributions should overlap as little as possible. Second, the correlation between graph properties should be small in order to avoid redundant information.

In this section we considered only the first moments (means) of the graph properties taken from an individual brain. However, as explained in section 3.4.4, we use the first four moments of the individual distributions for our statistical analysis. Since the p-value of the other moments is not calculated, its influence on the statistical analysis cannot be considered.

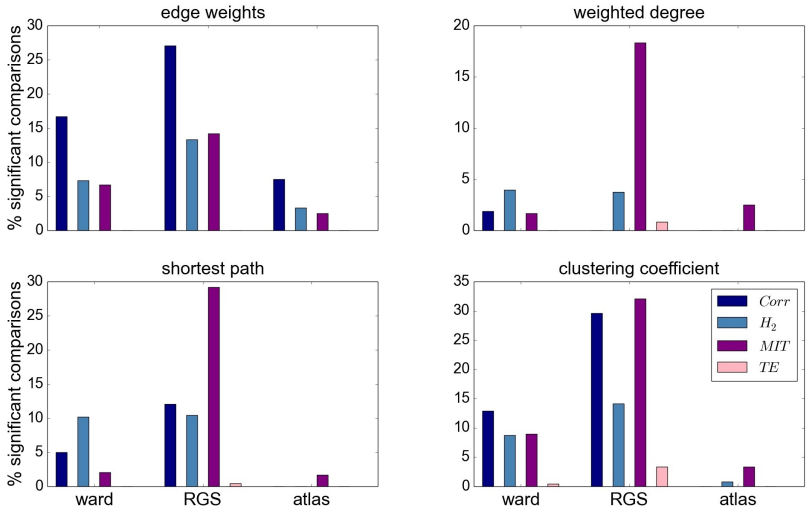


**Figure 3.6: Significant relationships in graph metrics between heath conditions dependent on edge definition methods.** Percentage of significant differences for each clustering technique (ward (left), RGS (middle), atlas (right)) for each class of edge definition method clustering method ($corr$ (dark blue) , $H_2$ (light blue), $MIT$ (purple), $TE$ (pink)) for averaged graph properties: edge weights (top left), weighted degree (top right), shortest path (bottom left) and clustering coefficient (bottom right). Fraction of significant differences are calculated for each health condition over all graphs constructed with the corresponding clustering and edge techniques including all variants in parameters, thresholds and rich-club sub-graphs. Significance is calculated as in the lower panel of Fig.3.4.

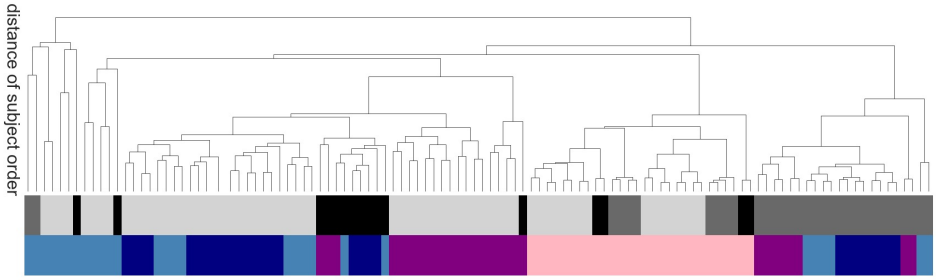


**Figure 3.7: Sensitivity of subject order to clustering and edge detection techniques** The dendrogram shows the distance of subject order, calculated by ordering all subjects according to their average graph edge weights and calculating the euclidean distance between the resulting rank arrays. For better legibility, instead of naming the dendritic leaves, of which every leaf corresponds to a particular combination of clustering and edge definition techniques, e.g ward2 $BTED2$, the top row of colors code for the class of cluster method: ward (light gray), RGS (dark gray) and atlas (black); and the bottom row codes for the class of edge definition method: $corr$ (dark blue), $H_2$ (light blue), $MIT$ (purple) and $TE$ (pink).

In order to evaluate the methods based on robustness due to methodical variation, we investigate how the order of subjects (all subjects independent of their health conditions are

ordered according their average value of a certain graph property) is affected by the exact realization of the graph construction methods. Graphs constructed by methods based on similar underlying features of the data will tend to show a systematic ordering of subjects, regardless of the absolute values of the calculated graph metrics. Fig.3.7 illustrates the commonalities and differences are illustrated with a dendrogram (see section 3.4.7.2) calculated on the Euclidean distance between the resulting ordered arrays of average graph weights. The continuous pink area show that graphs constructed using transfer entropy are most robust to the choice of clustering technique. Moreover, linear and non-linear correlations (dark and light blue) occupy contiguous blocks and so are most similar to each other. The leaves denoting atlas clustering (black) are rather spread out, indicating a high sensitivity of this method to the choice of edge definition.

In this section we have shown that the relationship of graph properties between health conditions strongly depends on the methods used for graph construction. For our data we find more significant mean differences for control-AD and control-MCI then for MCI-AD. With respect to clustering and edge definition methods, the most significant differences are found for RGS and Ward clustering, and for model-free edge definitions. These results show that conclusions on how graph properties change due to AD have to be drawn carefully, and ideally validated by other methods, as they can be highly sensitive to the methods used for graph construction.

### 3.2.3 Evaluation of graph construction methods based on negative surprise

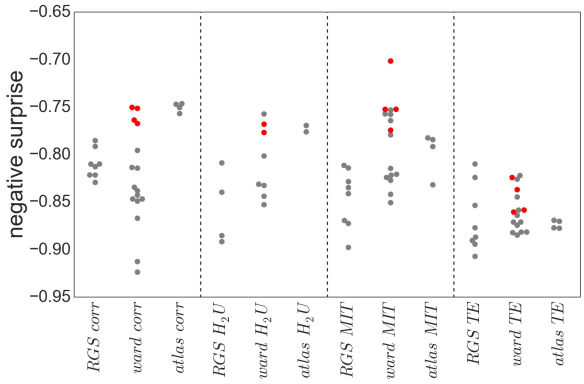


**Figure 3.8: Negative surprise of the different graph construction methods.** Negative surprises for all graph construction methods except $H_2D$. Each dot represents a specific node clustering (e.g. RGS1) and edge definition (e.g. $BCorrU1$). Dots are grouped together according to their main class (e.g. RGS *corr*). Red dots highlight the results of ward4 clustering. Negative surprise expected by chance is -1.1 .

Having examined the consequences of particular choices for clustering and edge-definition techniques in the previous sections, we now evaluate their combinations by considering

their ability to help a clinician to discriminate among patient groups. This discrimination is achieved by using the graph data within a statistical model, which specifies the likelihood of the graph data. The model is described in Sec. 3.4.6; the likelihood is a t distribution which depend on a set of parameters. In general, the kind of graph data – i.e. their construction method – and the statistical model with its parameters are interdependent: they cannot be freely varied separately. Therefore our evaluations of the predictive power of the various graph construction methods have to be understood with a caveat: they depend on our specific choice of statistical model.

To quantify the discriminating power for each graph construction combination, we use a metric based on the final probabilities for the correct health conditions known as the log-probability, or *negative surprise* (Bartlett, 1952; Good, 1956, 1957a): a sure event, i.e. with unit probability, has surprise equal to zero; whereas an impossible event, i.e. with zero probability, has surprise equal to infinity, reflecting the fact that its occurrence would be contrary to all our expectations. A high surprise (in absolute value) therefore signals a low predictive power of the data we are using. The expectation or average of the surprises is the Shannon entropy (Shannon, 1948; Bartlett, 1952; McCarthy, 1956; Bernardo, 1979; Jaynes, 2003, Sec. 11.3).

Another possibility, of a more decision-theoretical character, is to consider a metric based on the average utilities obtained with each particular graph-construction method. Given several possible courses of action (e.g., treat or dismiss) and their utilities or costs with respect to each health condition (e.g., treating an Alzheimer patient, dismissing a healthy patient, dismissing an Alzheimer patient, or treating a healthy one), the clinician should choose the action that maximizes the expected utility, the expectation being calculated from the final probabilities for the possible health conditions (Sox et al., 2013). This kind of metric therefore requires not only the final probabilities – which depend on the graph-construction method – but also a table of utilities.

Numerical tests show that the two kinds of metric yield similar results, at least for utility tables close to the identity (treating an ill patient and dismissing a healthy one have unit utility; the remaining combinations have zero utility). We therefore choose a metric based on the negative surprise, which is simpler and more intuitive than a utility metric.

In order to have an approximate idea of the relative predictive powers of the graph-construction methods we would like to use a statistical that can be kept the same, as much as possible, across different methods. For this reason we choose a model based on the working hypothesis of sufficiency of mean and correlations of past data, as explained in the Introduction. This model ignores any restricted range of variability of graph quantities (e.g. positive or bounded). As explained in Porta Mana et al. (2018), this choice is non-standard but does not entail contradictions. The model has some free parameters; their values reflect the fact that the units of measure for the graph quantities make the latter of order unity. This choice of a generic, common statistical model allows us to sidestep the demanding problem of tailoring it for the different graph quantities from our 850 graph-construction methods.

Fig. 3.8 show the obtained negative surprises for all combinations of graph construction methods except $H_2D$, which is left out due to inadequacy of the statistical model, resulting

in unrealistic values between $-1.26$ and $-0.66$ with a mean and standard deviation of $-0.94 \pm 0.19$.

The differences in negative surprise between the different graph construction method are in general small. The best results are obtained for ward4 clustering combined with mutual information ($MIT$) based edge definition. Across edge definition methods, linear correlation ($corr$) and mutual information give the best results and transfer entropy ($TE$) the worst. The rather poor performance of $TE$ edge definition is in line with the small number of significant differences found for this method (compare Fig.3.6). Comparing the different clustering methods, atlas and ward4 clustering give the best results, as long as the edge definition is not $TE$. These two clustering methods have in common a very small number of graph nodes and (correspondingly) the highest number of voxels per cluster (compare Fig.3.2).

As explained a few paragraphs above, the comparison of graph-construction methods can be affected by the statistical model and its parameters, especially for small datasets. As a complementary touchstone we compare the negative surprises with the classification performances of a support vector machine (SVM, section 3.4.7.1) based on the same graph constructions. In a clinical setting, a misclassification between control and AD has more severe consequences than between MCI and AD. To avoid introducing an asymmetric misclassification penalty, we perform the classification between pairs of classes only (control-AD, C-MCI, MCI-AD).

Figure 3.9 shows the relationship between the SVM performance (measured as proportion of correct classifications) against the negative surprise. As long as $TE$ edge definition is excluded, the two performance measures are positively correlated. In particular RGS clustering achieves low performance in both negative surprise and SVM classification. Furthermore, atlas clustering achieves a high classification performance across all edge definitions. The exact SVM classification results for each realization of graph construction method are depicted in Fig.S2 in the supplemental material.

Figure 3.10 demonstrates that thresholding graphs has only a minor effect on the negative surprise for small thresholds up to 0.2. No systematic relationship can be observed for the effect of larger thresholds; for example, increasing the threshold to $0.4$ causes a decrease in negative surprise for RGS clustering with linear correlations or mutual information, but an increase for atlas clustering with transfer entropy edge detection. Likewise, the creation of highly connected and rich-club sub-graphs typically decreases the negative surprise, but in some cases increases it (e.g. RGS $H_2U$). Overall the highest negative surprise ($-0.66$) is obtained for ward4 clustering combined with $BMITU1$ thresholded at $w_{\min} = 0.1$.

These results suggest that the best combination of graph construction techniques to use for this data set is the atlas-based or ward4 clustering combined with linear correlation methods or mutual information transfer. Thresholding the graph edges, which might reduce experimental noise and lowers computational complexity, has only little effect on the predictive power, as long as threshold values are small. Reducing the graphs complexity via larger thresholds or extracting the rich-club of the graph should be done with care, since the results can change in either directions. Although transfer entropy yields lower negative surprises then the model-free functional connectivity measures, we would not conclude that this edge definition performs worse in general, since it achieves high values in SVM classification. It is very likely

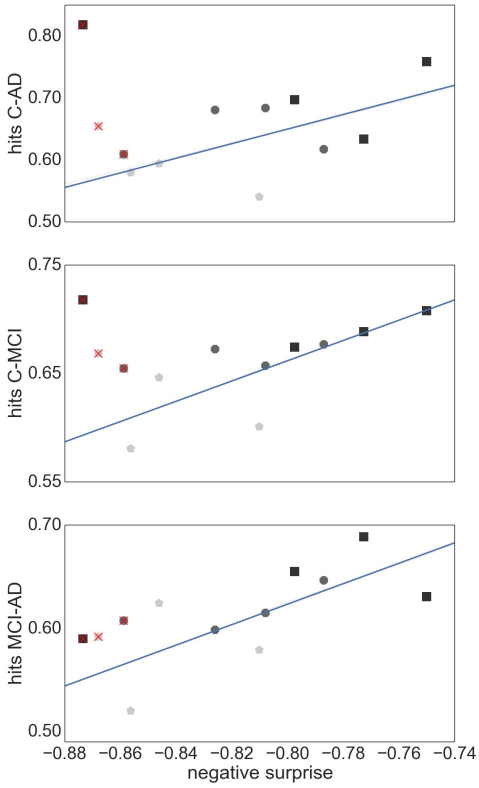


**Figure 3.9: Relationship between SVM classification performance and negative surprise.** the average SVM performance achieved by each combination of clustering method and edge definition with respect to each pair of health conditions: control-AD (top panel), control -MCI (middle panel) and MCI-AD (lower panel), is plotted against the negative surprise calculated for all health conditions. Each marker corresponds to the averaged performance across the parameter space of a specific clustering method (atlas (black squares), Ward (dark gray octagons), RGS (light gray pentagons)) and a specific edge defintion ($corr$, $H_2$, $MIT$, $TE$). The regression line is calculated for all points but $TE$ (superimposed red crosses). Pearson correlation coefficients $r$ of the datasets are $r = 0.59$ (top panel), $r = 0.77$ (middle panel), $r = 0.69$ (lower panel).

that our choice of statistical model is not ideal, and a more tailored choice would improve performance.

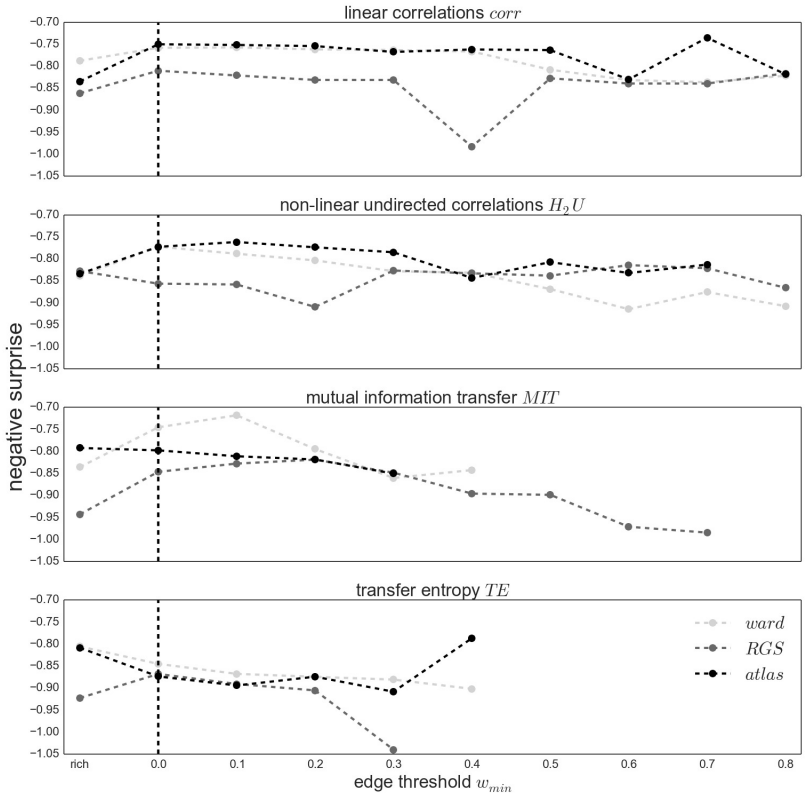


**Figure 3.10: Negative surprise in edge-thresholded graphs.** Negative surprise for different graph edge thresholds $w_{\min}$ ($w_{\min} = 0$ for complete graphs, stressed by a vertical dashed line) and richclub graphs (rich) across different edge definitions: $corr$ (first panel), $H_2U$ (second panel), $MIT$ (third panel), $TE$ (last panel) and different clustering methods ward (light gray), RGS (dark gray) and atlas (black). Every dot is the result of averaging across all possible parameters of a general graph construction method (for $w_{\min} = 0$ the average across all points of a swarm in Fig.3.8). Since some methods yield small edge weights, some graphs become unconnected for large $w_{\min}$ such that the statistical analysis is not conducted and no values (dots) are depicted here. Markers are connected for better visual comprehension.

## 3.3 Discussion

In this article we have compared different techniques for constructing and analyzing graphs. By applying a statistical model, we have demonstrated a principled method for choosing a combination of techniques for a given data set. By examining the varied outcomes of the techniques, we have shown how sensitive the results of graph theoretical analyses, such as

significant differences in mean properties, can be to the choice of clustering or edge definition technique.

With regards to the predictive power of the graph construction techniques, measured in terms of negative surprise, we find that Ward and atlas clustering yield the highest performance of the clustering techniques, and region growing and selection clustering (RGS) the lowest. In particular, the variant of Ward clustering that produces large brain clusters and small numbers of nodes (ward4) achieved the highest performance values. Analogously for the edge detection methods, we find better performance for the model-free methods (linear and non-linear correlations, mutual information transfer) than for the model-based method of transfer entropy. For this particular data set, a combination of ward4 clustering with mutual information derived edges achieves best results. Therefore, we would recommend this combination as the primary target for a more narrowly focussed investigation based on a larger data set.

The performances we obtain are above chance level but still far away from optimal prediction of the three health conditions. One reason for this sub-optimal prediction might lie in our choice of statistical model and its parameters. With our small data set (26 controls, 16 MCA, 14 AD) the model and its parameters have a high influence on the final probabilities, and thus on the performance (Porta Mana et al., 2018). We avoided tailoring the statistical model for the theoretic and practical reasons explained in Sec. 3.2.3. Even if the model is not tailored, the results are consistent with the classification performance of support vector machines (see Fig.3.9 and supplemental data Fig.S2), for the model-free edge definition techniques.

It remains unclear why Ward and atlas clustering are more successful than RGS, especially in combination with model-free edge definition. One possibility is that this is related to the large variability in graph sizes generated by RGS (Fig.3.2). In addition, the variance of weight distributions across subjects, and the variance of the cluster distances, are much larger in RGS then in ward clustering (Fig.3.3). This could be related to the variance in the number of nodes; however, choosing graphs similar in size causes even higher variances (Sec. 3.2.2). Therefore we assume that the number and connectivity of the small functional units extracted by RGS are highly variable across subjects. This variance might be even higher across subjects within a health condition than across health conditions, such that changes due to AD cannot be detected. This assumption might at first glance seem to contradict the high number of significant comparisons observed (Fig.3.5). However, we only calculate the significance level for the means of the distributions and not their entire shape. In addition, it is likely that some graph properties correlate with the graph size, and thus that apparent significant differences in graph properties are simply reflecting significant differences in numbers of nodes detected, and do not provide further information useful for classification or understanding the nature of the disease. Further investigation is needed on this matter.

The low negative surprise of transfer entropy ($TE$) compared with other model-free functional connectivity measures might have several reasons. The comparison of the negative surprise with the support vector machine classification suggests that a better choices of statistical model is possible: the classification results for $TE$ are similar to those of the model-free measures. In $TE$ the data of a certain time interval in the past is used in order to calculate how much the uncertainty of the future is reduced. Here we use the data of the last 15

sec. This time period might be poorly chosen, influencing the overall negative surprise. In addition $TE$ is more sensitive to short recording periods than other methods, which may well also result in a reduced performance (Pereda et al., 2005).

With regards to the robustness of the graph theoretical outcomes, we discovered that relationships between mean graph properties, such as closeness centrality, edge weight or clustering coefficient (Figs. 3.4 and 3.5) were sensitive to choice of clustering and edge definition techniques, to parameter choices for a given technique, and to the manner in which sub-graphs were defined (thresholding value and rich club). For most relationships between graph properties $X$, we could find significant ($p < 0.05$) differences in both directions, i.e. both $X_{\mathrm{AD}} > X_{\mathrm{C}}$ and $X_{\mathrm{AD}} < X_{\mathrm{C}}$, for specific choices of clustering and edge definition technique. This strongly suggests that a degree of suspicion should be applied to studies reporting such significant differences, especially if these results are argued to give insight into how a disease affects brain properties, unless the significance level is much more compelling or the reported differences can be validated with alternate methods.

We also investigated the sensitivity to method choice of the ordering of subjects according to a graph theoretic metric (Fig.3.7). In this analysis, transfer entropy was the most consistent. Nevertheless, the distribution of the negative surprises is as broad for transfer entropy as for other edge definitions (Fig.3.8). In general, the exact parameter selection within an edge definition method causes only slight changes in the negative surprise, more crucial is the exact realization of the clustering method: ward4 clustering generally achieves a better performance then ward3 clustering. These two variants differ only in the number of predefined clusters (see supplemental material Fig.S1). Applying a lower threshold $w_{\min}$ on the graph's edge weights has little effect on the negative surprise for all methods, as long as only small weights (up to $0.3$) are set to zero. Thresholding higher weights or extracting the graph's rich-club has unpredictable effects on the results, and so should be used with caution (Fig.3.10). Atlas clustering was least consistent in the subject ordering analysis, suggesting that although it may provide a good basis for a diagnostic tool, care should be taken in reporting discoveries of particular relationships in graph properties between health conditions, as these may well turn out to be critically dependent on the edge definition method used.

Due to the intense computational requirements of the survey performed in this article, we recognize that it would be advantageous develop heuristics for choosing between graph construction methods without performing the full calculation for each combination. Our results suggest that properties visible at the clustering stage, such as average heterogeneity, may give some indication of predictive performance: graph constructions that result in different degrees of heterogeneity between the health conditions seem to be more discriminable by the later steps of the calculation. More research is needed in this area, which is outside the scope of the current study. In addition, it is tempting to consider t-test results of the mean graph properties as a heuristic. Our results suggest that this approach is largely inadequate. It holds for edge definition via transfer entropy, which gives very few significant results and the negative surprise is rather small compared with the model-free edge definitions. Conversely, region growing clustering yields most significant differences but a generally poor negative surprise. This may be due to graph properties being highly correlated, and so not providing additional information to the statistical model. In addition we used the first four

moments (wherever possible) in our statistical model, rather than just the mean, which may also partially account for this apparent contradiction.

In addition to considering the predictive power and robustness of graph construction techniques, we can also evaluate them according to their practicality, i.e. speed of calculation and the extent to which they are easily available in established medical infrastructure and diagnostics. In general, applying graph theoretic measures to fMRI data for improving AD diagnosis makes sense, since MRI scans are already implemented in AD diagnostics for detecting structural changes such as hippocampal dystrophy caused by AD or AD-unrelated pathology (e.g. brain tumors). Softwares such as SPM (Tzourio-Mazoyer et al., 2002) and FSL (Jenkinson et al., 2012) are frequently used in medical research and mainly support clustering that is atlas and independent component analysis based. Ward clustering, which is the fastest of all these clustering methods, is a standard hierarchical clustering method and implemented in all standard programming softwares such as Python and Matlab. The region growing algorithm is not implemented in established softwares and is also computational very demanding. Given that it does not out-perform atlas or Ward clustering, we therefore do not recommend it. For edge definition and graph properties, several software packages are available based on Matlab (Kruschwitz et al., 2015; Wang et al., 2014) or Python[1], which provide a comprehensive range of edge definition and graph analysis methods.

In general we recommend using statistical models and not pure classifiers such as support vector machines as diagnostic tools, since statistical models calculate a probability of a diagnosis rather than assign a classification, i.e. 'Given the fMRI scan, person x has a $80\%$ probability of having Alzheimer's disease', rather than 'Given the fMRI scan, person x has Alzheimer's disease'. Probabilities can be easily combined with other probabilities of other diagnostic tests (Porta Mana et al., 2018) such as cognitive assessment, amyloid beta and tau protein occurrence in cerebrospinal fluid, blood tests, and structural MRI[2](Johnson et al., 2012). This allows the medical doctor to conclude, for example: 'Given the results of the cognitive test *and* cerebrospinal fluid analysis *and* structural and functional MRI scan, person x has a $95\%$ probability of having Alzheimer's disease'. After the estimation of the probability for a disease, she has to decide on a treatment, also taking into consideration such factors as 'how harmful would the treatment be for a healthy person', which can be expressed in a utility function (Porta Mana et al., 2018). In addition, the statistical model used in this work allows an estimation of how much the model can be trusted, and therefore evaluate whether the sample size is sufficiently large (Porta Mana et al., 2018).

### Relationship to previous studies

Studies focusing on the graph properties extracted from resting-state fMRI in AD and its pre-stages generally have one of two aims. The first aim is to identify significant differences in the graph properties between health conditions, and to use these to gain insight into the effects of AD on the physical brain and its cognitive processes. These studies complement the picture revealed by investigations based on structural MRI and functional changes on

[1] `https://github.com/dpisner453/PyNets`
[2] `https://www.alz.org/research/diagnostic_criteria/`

the basis of EEG and MEG recordings. Typically a variety of graph properties (e.g. nodal degree, clustering coefficient, averaged shortest path, local efficiency, betweenness centrality, global efficiency, small worldness) are calculated, and used to motivate an account of how disease-related modifications to these properties result in a reduced capacity to transfer and process information.

However, such studies reveal entirely contradictory results. For example, the value of the clustering coefficient in AD with respect to controls has been reported to be increased, unchanged, and decreased, respectively (Zhao et al., 2012; Sanz-Arigita et al., 2010; Supekar et al., 2008). Analogous contradictions have been found for the comparative length of the shortest path (Sanz-Arigita et al., 2010; Supekar et al., 2008; Zhao et al., 2012). These contradictions could be caused by methodological differences or by not separating the different states of AD. Our results show ample evidence that the precise choice of graph construction techniques can easily account for contradictory findings, even for atlas based clustering, in which the number and size of clusters is held constant across all subjects (Fig.3.5). Evidence that the separation of different AD stages is relevant was provided by Kim et al. (2015), who demonstrated a non-monotonic behavior of global efficiency, local efficiency and betweenness centrality across different stages of AD and MCI. In our study, we could reproduce the pattern of increase and decrease of closeness centrality across conditions (Fig.3.4). However, we also demonstrate that the same analysis based on the rich-club sub-graph yields a different pattern, and that contradictory (but significant) results can be obtained for the same graph construction techniques with different choices of threshold. We thus conclude that differences in graph properties between health conditions are currently ill-suited to provide an account of disease mechanisms in AD, unless either: 1) a specific method of graph construction can be shown to be more representative of the underlying connectivity than other methods, 2) the differences can be shown to be robust to choice of graph construction, 3) the differences can be validated by another analytical approach, or 4) the significance level is shown to be substantially more persuasive than $p < 0.05$.

The second category of studies use graph theoretical information as input for machine learning algorithms to classify the health conditions of the subjects. Note that for this purpose it is irrelevant if a difference between health conditions is not robust to method choice, as the goal is not to understand the effects of the disease but to robustly distinguish between conditions. Recent studies have reached very high performance: $100\ \%$ accuracy in discriminating AD and control Khazaee et al. (2015), and $93\ \%$ for AD, MCI and control classification (Khazaee et al., 2017). In the latter work they extract more than two dozen local and global graph properties, resulting in roughly $3000$ features, since each of the local properties is calculated for all brain areas. Only a small subset of features is then used for classification, e.g. in-degree of the left middle temporal gyrus. They found that the classification power of local graph measures is larger than that of the global ones. Local changes in graph properties that do not propagate to global mean values have also been reported for area specific (frontal cortices, parietal and occipital regions) synchronization levels. (Sanz-Arigita et al., 2010)

In this work we do not compare node-specific graph properties, because Ward and RGS clustering do not result in the same spatial location of clusters across subjects. Instead, we consider, wherever possible, the first four moments of the entire distributions of graph

properties. This is more information than typically used for global measures, where often only the first moment (the mean) of a graph property distribution is taken into consideration. Nevertheless, it is still possible that considering single nodes, of which some may be more damaged by AD than others, could yield a better diagnostic performance. that single nodes, as they are more attacked by AD then others, give a better diagnostic performance. This requires further study in a survey considering only atlas based clustering. Again, this is out of scope of the current study, but we remark that the statistical model methodology we employ here would be equally applicable to such an investigation. The advantage of taking the entire distribution lies in the possibility of using purely data driven clustering algorithms (e.g. Ward clustering) that can be substantially faster than atlas based clustering, since they do not depend on a time and memory consuming registration of the individual brain image to standard space. In addition, the global distribution is more likely to be more robust against brain morphologic abnormalities such as brain tumors or brain shrinkage, and is more stable across recording sessions Wang et al. (2014); Telesford et al. (2010). Finally, a short recording time might be expected to have a weaker influence on entire graph property distributions then on single nodes. Thus we conclude that global measures are preferable, if a good diagnostic performance can be reached. Although the goal of this work was not classification, we note that we obtain up to $80 - 90\%$ correct classification using an off-the-shelf support vector machine on leave-one-out subsets of our data for pairwise (C-AD, C-MCI, AD-MCI) comparisons.Whether global measures can reach the impressive performance shown by Khazaee et al. (2017) can only be investigated on a sufficiently large data set, ideally with several hundred participants.

### Limitations of this study

In each step of the graph construction and analysis pipeline (Fig.3.1) we set limits to the endless space of possible methods and their corresponding parameters. Here we will shortly summarize the reasons motivating the selection of the methods examined here and the exclusion of others, given the constraint of limited computational and temporal resources. As a general principle, we aimed to include the most commonly used method(s) and additional methods that we found to be reasonable, even if they are not currently frequently used.

Starting with the fMRI pre-processing, we had to decide whether to include global signal regression. The global signal (the average activity across all brain voxels) is assumed to originate partly from vascular and respiratory processes that not represent neuronal activity. However, there is also evidence that it contains neuronal-signaling based components, since it is negatively correlated with the EEG signal and strongly correlated with the activity of the largest network in the brain (the default mode network, which plays a major role in rest state activity) when noise levels are low (Murphy & Fox, 2017). Without global signal regression, the Pearson correlation distribution derived from the signal of all voxels, or the average activity of clustered voxels, is biased to the left such that negative values are rare and small. The correction for the global signal centers this distribution, such that negative values are much more prominent. This also changes the properties of the graphs extracted from such data, for example an increase in modularity combined with fewer unconnected nodes has been reported (Schwarz & McGonigle, 2011; Hayasaka, 2013).

Speaking against global signal regression is the finding that correction for white matter, CSF and motions yield the most stable graph properties across sessions compared with additional applied global regression (Schwarz & McGonigle, 2011). In diagnostics it is important to have only small variance in the outcome across different sessions if the health condition of a subject is stable, such that small changes that indicate a worsening of the health condition can be rapidly detected. Moreover, we define the edges of our graphs as the absolute values of the functional connectivity values. As the negative part of the correlation distribution is small without global regression, different possible treatment of negative correlations (taking the absolute values or setting them to zero) should have only a small influence on the resulting graph properties, at least when the underlying functional connectivity are based on correlations. Consequently, we elect not to include global signal regression on our pipeline.

In the clustering step, the most commonly used method is to define clusters based on cortical regions defined by a brain atlas. We supplemented this with two data-driven clustering approaches: Ward clustering and RGS clustering. We selected Ward clustering, as it has been shown to perform better than alternative hierarchical clustering methods with respect to reproducibility and accuracy (Thirion et al., 2014). RGS, a method derived from image processing (Lu et al., 2003), was selected because we could adjust the method to produce functionally homogeneous clusters. In this formulation, the only free parameter of the algorithm is the minimal cluster size. For both the data-driven methods, we selected parameters such that graphs did not exceed a maximal size of 1500 nodes, due to computational limitations. We excluded clustering based on independent component analysis, because of its laborious implementation and the requirement for domain expertise to distinguish noise from activity-related components. We also excluded all clustering algorithms that do not take functional consistency into account, e.g. dividing the voxels into cuboid patches, as has been proposed for structural data (Amoroso et al., 2017).

With regards to methods for edge definition, we limit our survey to functional connectivity measures that act in the time domain and not in the frequency domain, thus omitting frequency based wavelet analysis (Supekar et al., 2008), synchronization likelihood (Sanz-Arigita et al., 2010) and coherence (Wang et al., 2014). The most commonly used and simplest functional connectivity measure is the Pearson correlation coefficient (e.g. Zhao et al., 2012), which we name $BCorrU$ in our work. We also test two additional model-free and one model-based method. A further model-based method based on Granger causality was excluded because it is too computationally expensive for larger graphs (Wang et al., 2014).

A thresholding operation is often applied to graphs extracted from fMRI, setting all values below $w_{\min}$ to zero. The aim of this step is to reduce experimental noise, which mainly manifests in the weaker edges, and to make the computation of graph properties computationally less demanding (Bordier et al., 2017). The threshold $w_{\min}$ can be defined in several ways: It can be set arbitrarily , without satisfying a certain demand, or such that certain properties of the graphs are preserved, e.g. average number of edges per vertex (Sanz-Arigita et al., 2010), node density (Zhao et al., 2012), small word behavior (Bassett et al., 2008) or a fixed cluster coefficient. Alternatively, it can be set such that information on the network's community structure is maximized e.g (Bordier et al., 2017). In a variant of the thresholding approach, it has been proposed to transform the edge weights by applying a power law (Schwarz & McGonigle, 2011). In this study, for the sake of simplicity, we examine graph properties as

a function of $w_{\min}$ without targeting any specific value of a graph property for simplicity. Potentially, our results would reveal a different picture if $w_{\min}$ was optimized for each subject to attain, for example, a specific average nodal degree. However, comparison of these two different thresholding mechanisms resulted in no major difference in the relationships of graph properties between the control and AD groups (Sanz-Arigita et al., 2010).

We do not binarize our graphs (setting all values below $w_{\min}$ to zero and those above it to one ) as is frequently done (Zhao et al., 2012, e.g.), as this leads to a loss of information, and moreover some distributions of graph properties distribution would become discrete (e.g. only ones and zeros for edge weights distribution), such that higher moments would be uninformative. The disadvantage of using weighted graphs lies in the limitation of possible graph properties. Most graph properties are well-defined for binary graphs and have been partly extended to weighted graphs. Here, we calculate the (normalized) weighted degree, shortest path, closeness centrality, clustering coefficient, and the modularity of the undirected graphs. We only investigate the most commonly used metrics and do not include more complex method such as the minimal spanning tree (Koray, 2011).

In addition to the restrictions of scope with regards to the examined techniques, a clear limitation of this study is the small data set. As our aim here is primarily to provide a methodology for evaluating and comparing analysis methods, rather than to draw conclusions on the effect of Alzheimer's disease on the graph properties of the cortex, a small data set is less problematic. Indeed, for the explorative survey carried out here, a large data set would have been prohibitively expensive with respect to computational resources. Moreover, many studies applying graph analysis to fMRI data are based on similarly sized data sets, which highlights the importance of raising awareness about the methodological artefacts we have identified.

The results of our survey indicate which combinations of methods are promising in view of Alzheimer diagnosis and which should be investigated further in future studies based on larger data sets. Naturally, such a study could yield some quantitatively different results, particularly with regard to the classification performance. Nonetheless, we would like to summarize some conclusions of the work that are unlikely to change with a larger data set. First, our results show that different combinations of methods can lead to contradictory findings with regard to significant differences in mean properties (Sec.3.2.2). This effect is unlikely to be resolved by a larger sample size. Second, methods showing good robustness with respect to parameter choice for a small sample size (e.g. $TE$ edge definition, see Fig.3.7), are likely to remain robust with increasing sample size. Likewise, there is no reason to assume that methods performing well in all circumstances for the small data set, e.g. MIT clustering combined with $corr$ edge definition (Sec.3.2.3), would perform worse for larger data sets. Finally, we assert that thresholding the graphs of a large data set with a small $w_{\min}$ (as shown in Sec.3.2.3) would similarly not result in a sudden jump in negative surprise.

### Application of approach to other analysis techniques

We have demonstrated a systematic, quantitative approach for comparing and evaluating sequences of algorithms that result in classification of fMRI data based on the first four moments of simple graph theoretic metrics defined on the whole graph. However, the approach we present is equally well suited for assessing pipelines based on other metrics. as we briefly outline in the following.

One possibility is to consider the graph properties of individual nodes, as these have been shown to be very informative (Dillen et al., 2017; Khazaee et al., 2015; Wang et al., 2016; Xia et al., 2014). This entails the use of atlas based clustering. We speculate that a global analysis of graph properties would be both faster and more robust to brain abnormalities and short recording times, and so would be the preferable approach if equivalent performance levels can be attained.

A second possibility is to extend our approach to a hierarchical analysis. This could potentially be of great use, as previous studies based on PET imaging have suggested that in Alzheimer's disease, long range connections become weaker but local clustering increases (Pagani et al., 2016, 2017). These alterations would not be observable using the graph analyses so far considered, although we have taken the first step by calculating the modularity, which compares the ideal dissection of the given graph into modules with that of a random graph with similar edge weights.

To capture the graph meta-structures it is necessary to cluster graph nodes into modules, or sub-graphs. Modules can be defined either purely functionally, such that each node (ideally) has the strongest connections to the nodes in its own cluster, and the weakest connections to nodes of other clusters, or based on anatomic structures, such that nodes in a cluster are part of large, anatomo-functionally similar brain areas. Analogous to the variety of methods for spatial clustering and edge definition investigated in this study, there are many techniques used to cluster nodes into modules (e.g. k-clustering, hierarchical clustering and spectral clustering, for review see Schaeffer (2007) or anatomo-functional clustering, see Pagani et al. (2016)) , and likewise multiple options for analysing the characteristics of the resulting modular structure (e.g. module degree or participation coefficient; see Guimerá & Nunes Amaral (2005)). Such a comprehensive study is outside the scope of the current work, but could well provide great insight into health condition related alterations in the global network structure of the brain.

### Conclusions

In order to achieve a robust and successful Alzheimer's disease diagnosis based on graphs extracted from fMRI data, we recommend clustering that results in rather small graphs with large clusters. Ward clustering, in which the number of clusters can be predefined, is fast, but requires programming knowledge to implement it. Atlas clustering is well established standard fMRI analysis software applications, but it is slow and might be affected

by morphologic abnormalities in the brain, such as atrophy which is a common symptom of AD.

Edge weights should be calculated based on correlations or mutually information transfer, especially if a focus of the study is uncovering significant differences in mean graph properties between health conditions. We emphasize that the existence, magnitude *and direction* of such significant differences can be very sensitive to the methods chosen, and the parameterization of those methods, and so such findings should be reported with care, especially if a biological interpretation of said findings is claimed. Transfer entropy rarely gives significant results, but is more robust towards parameter changes in the algorithm and different clustering algorithms. Finding appropriate statistical models may be an additional challenge for this method.

Weak thresholding may be used for complexity reduction as it has little effect on performance. Applying a higher threshold or extracting the rich-club sub-graph (10 % of the nodes with highest degree) causes unsystematic changes in the negative surprise and should therefore be used with caution, and validated against the full graph.

In summary, our quantitative evaluation and comparison of graph construction and analysis methods provides insight into how contradicting results come about in studies of graph properties of fMRI data, and identifies a number of potential methodological artefacts. Moreover, it provides a blueprint for establishing appropriate analysis pipelines, and serves as a well-founded starting point for future research on larger data sets.

## 3.4 Methods

### 3.4.1 Data aquisition

The recruitment and neuropsychological assesment of the study participants is given in Dillen et al. (2017). Demographic information is given in Table 3.2.

| | **Controls** | **MCI** | **AD** |
|---|---|---|---|
| number | 26 | 16 | 14 |
| Age | 62.38 [50, 73] | 70 [55, 78] | 71 [61, 78] |
| Sex | 10 f, 16 m | 7 f, 9 m | 7 f, 7 m |
| Years of Education | 15.3 [8, 25] | 12.75 [8, 21] | 12.83 [7, 18] |

**Table 3.2: Demographic information of participants** Average and minimal and maximal values [min, max] are given for age and years of education; female (f), male(m).

Anatomical MRI and rfMRI images were obtained from a 3T MR-Brain-PET scanner (Siemens, Erlangen, Germany) in the Memory Clinic Cologne Juelich. The parameters for the single-shot echo planar imaging sequence of the functional (T2* weighted) image are the following:

TR=3000 ms, TE=30 ms, FA=90°, FOV=200 × 200 $\mathrm{mm}^2$, matrix=80 × 80, voxel resolution = 2.5 × 2.5 × 2.8, 50 oblique slices parallel to the infra-supratentorial line, gap=0.28 mm, interleaved, scan time = 7 min. Parameters of the high-resolution T1-weighted structural image based on a magnetization-prepared rapid gradient echo sequence: TR= 2250 ms, TE=3.03 ms, FA=9°, FOV=256 × 256 $\mathrm{mm}^2$, matrix=256 × 256, voxel resolution = 1 mm isotropic, 176 sagittal slices, no gap, interleaved, scan time=314 sec. For more detail see (Dillen et al., 2017).

### 3.4.2 Preprocessing of fMRI-data and extraction of cortical data

Image preprocessing is accomplished using FMRIB's Software Library tools (Jenkinson et al., 2012; Woolrich et al., 2009). We carry out the following steps for the structural T1-weighted image: skull-stripping (Smith, 2002b) with bias field correction (Keihaninejad et al., 2010; Leung et al., 2011; Popescu et al., 2012) and for the functional T2-weighted image: discarding the first 10 volumes (out of 140 each taking after 3 sec), motion correction (Beckmann & Smith, 2004), spatial smoothing using a 4 mm full width at half maximum Gaussian kernel, high-pass temporal filtering at 0.02 $Hz$ and a six-parameter, rigid-body linear transformation procedure in MCFLIRT (Jenkinson et al., 2002). More details can be found at (Dillen et al., 2017), where the same preprocessing is applied. In addition we carry out white matter and cerebrospinal fluid regression (FSL regfilt, MELODIC) to the functional image in order to reduce noise.

In order to extract only cortical voxels from the entire brain fMRI image, as needed for the data-driven clustering described in the next section, we first register cortical regions (frontal- , occipital -, temporal- and insular-cortex) defined in the MNI structural atlas (Collins et al., 1995) to the structural and then to the functional space. For this registration we apply the transformation matrix obtained from registering the entire standard brain first to the individual structural brain (linear registration with FSL/FLIRT; Jenkinson & Smith 2001; Jenkinson et al. 2002) ) and then to the functional space (non-linear registration with Advanced Normalization Tools (ANTs; Avants et al. 2011). In order to extract only gray matter tissue, we apply the gray matter image of the structural space (segmentation with FSL-FAST (Zhang et al., 2001)) registered to functional space as described above, as a mask to the to the functional image.

### 3.4.3 Data-driven and Atlas based clustering of cortical voxels

In order to construct graphs we cluster cortical voxels into regions using three different methods. Two of these methods, the Ward clustering and the region growing and selection algorithm (RGS) are data driven such that only neighboring voxels with similar activity are combined into a single region. For these algorithms the number of regions per brain and the participating voxels in a region can differ for each individual and strongly depend on predefined algorithm-specific parameters. The atlas-based cluster algorithm on the other hand produces the same number of clusters and a constant number of voxels per region across individuals because the individual brains are mapped onto a standard brain

#### 3.4.3.1 Atlas-based clustering

For each subject we linearly register the rfMRI image first to the structural, skull-removed image (image segmentation for skull removing with SPM8, Wellcome Department of Cognitive Neurology, London, UKFSL; linear registration with FSL/FLIRT; Jenkinson & Smith 2001; Jenkinson et al. 2002) and then, through a non-linear mapping, to the MNI standard brai (non-linear registration with Advanced Normalization Tools (ANTs; Avants et al. 2011); MNI 152 standard brain, non-linear 6th generation (Grabner et al., 2006)). Regions of interest (ROIs) of the resulting functional image in standard space are extracted such that they match the 94 regions identified by the Oxford lateral cortical atlas (regions have a probability above $50\%$ )(Desikan et al., 2006). A demonstration of how the brain is clustered according to the brain areas given by the atlas is given in the first panel of Fig.3.12.

#### 3.4.3.2 Ward clustering

Ward clustering (python: sklearn.cluster.AgglomerativeClustering, Pedregosa et al. (2011)) is a data driven clustering algorithm, which is initiated by defining each voxel as a cluster and then, in each iteration step, merge the two neighboring clusters (even of different sizes) that after merging show minimal intra-cluster variance compared with all other possible variations of combining two adjacent clusters. In this way, the number of clusters is reduced by one in each iteration step. In our case the clustering stops after $k$-clusters (table 3.3) are formed. Afterwards, we discard away all clusters that comprise of less then $p$-voxels (table 3.3). An example of the outcome of ward clustering algorithm is depicted in the second panel of Fig.3.12.

#### 3.4.3.3 Region growing and selection

The region growing and selection algorithm is a modified version of the algorithm described in (Lu et al., 2003). Region growing implies that each voxel serves as an initial seed (center) and neighboring voxels are added iteratively if they fulfill a certain growing criteria. (Fig.3.11 a) The condition (Lu et al., 2003) proposed for adding a voxel to a region is based on the Pearson correlation coefficient $R$ between the averaged time-varying signals of the pre-merged region and the signal of the voxel to be tested. If this correlation is higher then a pre-defined threshold $T$ (table3.3), the voxel is merged to the region. We exacerbate the growing criteria by imposing a second condition that allow the merging of voxels only if in addition to exceeding the correlation threshold, the resulting cluster is also functionally homogeneous. Here, functional homogeneity means that the time-varying signals of all voxels can be expressed as instances of a single signal with varying levels of noise. The number of independent signals in a cluster can be estimated by the spatial functional heterogeneity $h$ (Marrelec & Fransson, 2011):

$$h = n_0 + \frac{e_{n_0} - b_{n_0}}{(e_{n_0} - e_{n_0+1}) - (b_{n_0} - b_{n_0+1})} \tag{3.1}$$

, where $e_n$ are the Eigenvalues of the $N$x$N$ covariance matrix of all $N$ time varing signals in a cluster that exceed the Eigenvalues generated by the broken-stick model $b_n$ such that $e_n > b_n = \sum_{i=n}^{N} 1/i$ . The index $n_0$ accounts for the smallest Eigenvalues that fulfill this inequality equation such that $e_{n_0} > b_{n_0}$ and $e_{n_{0+1}} < b_{n_{0+1}}$. A value of $h = 1$ indicates a homogeneous cluster.

The region selection algorithm iteratively selects the largest region and delete all clusters that have their centers in that region, excluding the possibility that centers overlap with other regions. Nevertheless clusters can still overlap (Fig.3.11 b). Applying this framework does not guarantee that clusters remain spatially connected after deleting regions with overlapping centers. Nevertheless a check for spatial consistency reveal that only a marginal percent of the clusters are disrupted in that way. Finally, we took only the clusters that comprised a minimum number of voxels $p$ (table3.3). How the outcome of RGS looks like is demonstrated in the last panel of Fig.3.12.

| method | minimal number of voxels per cluster $p$ | number of clusters $k$ | threshold $T$of Pearson correlation coefficient | |
|---|---|---|---|---|
| ward1 | 10 | 5000 | - | |
| ward2 | 25 | 5000 | - | |
| ward3 | 10 | 2000 | - | |
| ward4 | 25 | 2000 | - | |
| RGS1 | 55 | - | 0.75 | |
| RGS2 | 50 | - | 0.75 | |
| atlas | - | - | - | |

**Table 3.3: Parameters used for the different clustering algorithms.** Abbreviations: ward clustering (ward), region growing and selection (RGS), atlas-based clustering (atlas).

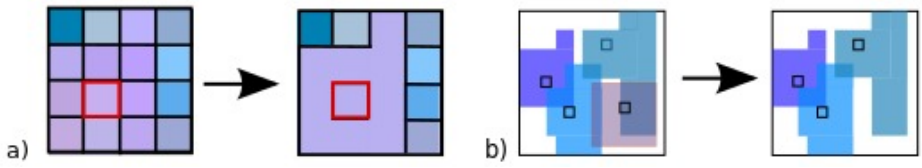


**Figure 3.11: Region growing and selection algorithm** a) Region growing, left: each voxel (colored squares) serves as center for a cluster, right: example of a growing region (purple), only adjacent voxels that fulfill the fusion criteria are added to the growing cluster. b) Region selection. Small regions (pink) with centers overlapping with larger regions (green) get deleted (from left to right) in a iterative manner. Remaining regions can still overlap as long as their centers do not cover other regions. This illustration is in 2D for simplicity, the algorithm used for fMRI data acts in 3D following the same rules.

### 3.4.4 Edge definition

A graph consists of nodes (vertices) that are connected through edges, that might be weighted or binary and directed or undirected. We construct individual brain graphs by defining nodes that represent clusters as described in 3.4.3, such that the mean activity of a cluster becomes an node attribute. We presume that all graphs are fully connected and edge weights are defined in terms of functional connectivity. Since functional connectivity can be calculated in several ways, we apply a range of different connectivity measures. In (Wang et al., 2014) many such methods are evaluated, taking the structural connectivity of a toy model as reference. As a starting point, for each proposed category of functional connectivity, measured in time, we select the analysis measurement that captures structural connectivity best. We follow this strategy for all proposed measurement categories in(Wang et al., 2014) , leaving out only Granger causality measures , due to limited computational resources. We thus use linear and non-linear correlation ($corr$ and $H_2$ ) and mutual information transfer ($MIT$) for the model-free category and transfer entropy ($TE$) for the model-based category. In all groups the bivariate methods perform better then the partial ones. In conclusion we select for each of the families the bivariate implementation that can be both directed and undirected. For consistency we use the same abbreviations for the different methods as in (Wang et al., 2014) and the same Matlab toolbox Mulan[3] which they made public. Here we provide only a short description of the applied methods and more details can be inferred from (Wang et al., 2014) .

Linear correlation ($corr$) are measured based on the Pearson correlation coefficient (Rodgers & Nicewander, 1988) in a pair-wise manner. For directed connectivity ($BCorrD$) delays of up to 5 time steps (table 3.4) are considered and the largest connectivity values is selected. We do not take into account time lags for undirected correlation ($BCorrU$).

Non-linear correlations ($H^2$) are based on piece-wise linear correlations of two time signals on which the non-linear curve is fitted (da Silva et al., 1989). Bivariate directed ($BH^2D$) and bivariate undirected ($BH^2U$) are defined as above for linear correlations.

Mutual information indicates how much information is shared between two time varying signals by means of Shannon entropy (Grassberger et al., 1991). For $BMITD1$ individual histograms of two time series are contrasted to the joint histogram across different time delays. No delays are taken into account in $BMITU$.

Transfer entropy (Schreiber, 2000) describes how far in the past, the activity of a node can reduce the uncertainty of the future activity of another node for which also the past activity is considered. Bivariate directed ($BTED$, (Chicharro, 2011) and bivariate undirected ($BTEU$) are defined as above for linear correlations.

All methods were tested for a window size that comprises the whole time range (130 time points/6.5 min) and for a sliding window of 50 time points (2.5min) with an overlap of 5 time points (0.25 min) (see table 3.4). If the methods revealed negative weights, the absolute value was considered . The resulting graphs are directed or undirected weighted graphs with

[3] https://github.com/HuifangWang/MULAN

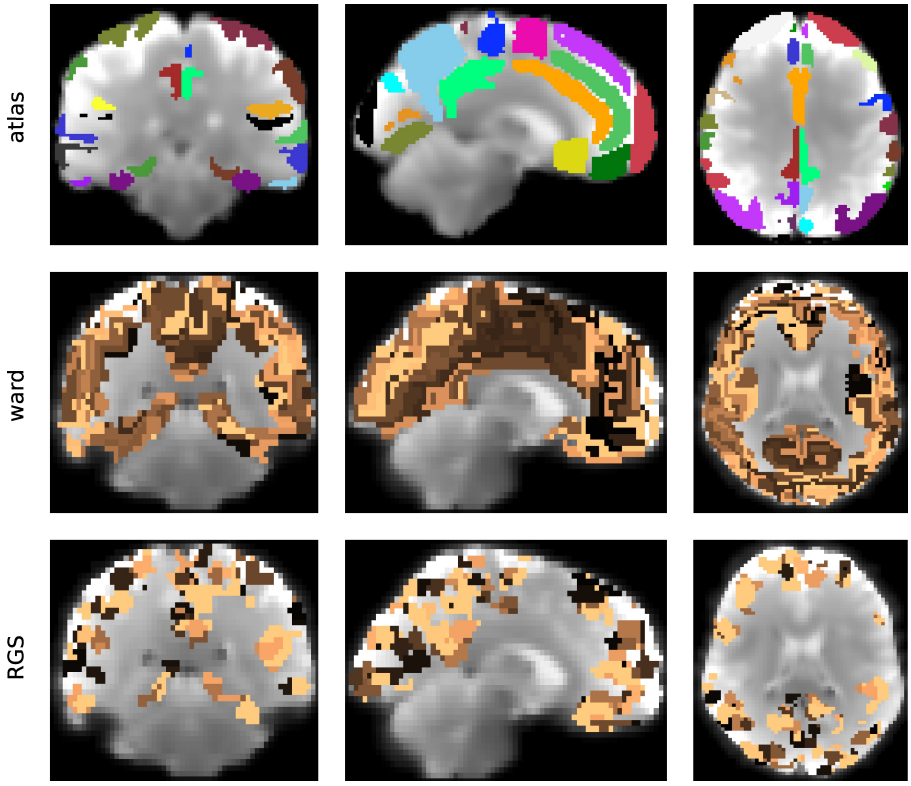


**Figure 3.12: Clustering of the cortical functional image.** Illustrated are the clustering outcome of the atlas (upper panels) and the ward clustering (ward4, middle panels) and RGS (RGS1, lower panels) algorithms for frontal, sagital and horizontal brain sections (from left to right) of a randomly chosen healthy individual. Individual clusters are depicted by a randomly chosen individual color, for clustering parameters see table 3.3.

values between zero and one for all methods except non-linear correlations, where values can exceed one.

Many studies transfer weighted graphs into binary ones by setting all values below a threshold $w_{min}$ to zero and above to one e.g. (Zhao et al., 2012). Following this strategy we also investigate the effect of setting all weights below $w_{min}$ to zero but leaving higher weights unchanged. As far as the remaining graphs are still connected (left panels in Fig.3.13 ) and single nodes are not disconnected from the network (right panels in Fig.3.13) we study the disease diagnosis capacity for $w_{min} \in \{0.1, 0.2, ...0.7, 0.8\}$. In addition we extract the rich club of the graphs. The rich club is a subgraph that comprises the nodes that are most strongly connected to the network. In this work we define the rich club as the 10% of the nodes with highest degree.

| method | window size | window overlap | number of bins | max. delay |
|---|---|---|---|---|
| $BcorrU1$ | 130 | - | - | - |
| $BcorrU2$ | 50 | 0.2 | - | - |
| $BcorrD1$ | 130 | - | - | 5 |
| $BcorrD2$ | 50 | 0.2 | - | 5 |
| $BH2U1$ | 130 | - | 10 | - |
| $BH2U2$ | 50 | 0.2 | 10 | - |
| $BH2D1$ | 130 | - | 10 | 5 |
| $BH2D2$ | 50 | 0.2 | 10 | 5 |
| $BMITU1$ | 130 | - | 5 | - |
| $BMITU2$ | 50 | 0.2 | 5 | - |
| $BMITD1$ | 130 | - | 5 | 5 |
| $BMITD2$ | 50 | 0.2 | 5 | 5 |
| $BTEU1$ | 130 | - | - | 5 |
| $BTEU2$ | 50 | 0.2 | - | 5 |
| $BTED1$ | 130 | - | - | 5 |
| $BTED2$ | 50 | 0.2 | - | 5 |

**Table 3.4: Parameters of the different functional connectivity measures** abbreviations: Bivariate ($B$), undirected ($U$), directed ($D$), linear correlation ( $corr$), non-linear correlation ($H_2U$) mutual information entropy ($MIT$), transfer entropy ($TE$)

### 3.4.5 Graph properties

This section describes the different graph properties that are either characteristics of single nodes (weighted degree, closeness centrality, cluster coefficient), of pairs of nodes (shortest path) or of the entire network (modularity). In the first two cases we get a range of values for each graph. Since we do not know, which are the important features of the resulting distributions, we decide to take the first four moments for our statistical analysis. Because graphs based on data-driven clustering results in different number of nodes and the calculated graph properties might be dependent on the number of nodes, we also include the number of nodes in the subsequent analysis (Sec. 3.4.6).

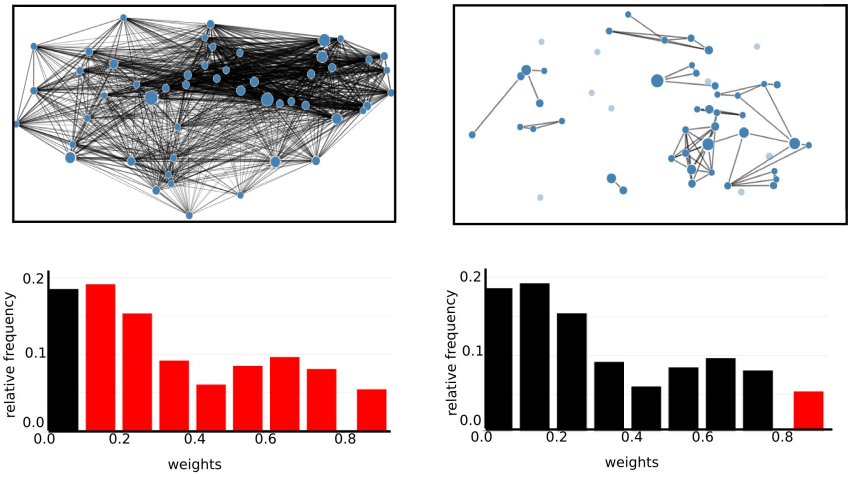


**Figure 3.13: High thresholds on graph edges cause the graphs to dissociate.** Upper panel: Illustration of graph with edge weights larger then 0.1 ($w_{\min} > 0.1$, left) and larger then 0.9 ($w_{\min} > 0.9$, right). In the according weight histograms (lower panels) the red bars correspond to the edges drawn in the upper graph. Edges corresponding to the black bars are not shown.

### Weighted degree

The weighted degree $deg_w$ describes how strong a node is connected to all other vertices of the network, obeying the equation:

$$deg_w(v) = \sum_{u \in V \setminus \{u\}} w_{uv} \tag{3.2}$$

where $w_{uv}$ is the weight on the edge between nodes $u$ and $v$ of all nodes $V$ in the graph. This definition implies a high dependency of the weighted degree on the number of nodes in a graph. To address this problem, we normalize the weighted degree

$$deg_n(v) = \frac{deg_w(v)}{deg(v) \cdot w_{max}} \tag{3.3}$$

with $w_{max}$ being the maximal weight of the graph. The resulting values are between 0 and 1.

### Shortest path and closeness centrality

The shortest path $dist_w(u, v)$ between a pair of nodes $u$ and $v$ describes the path that minimizes the sum of the weights of its participating edges. A small shortest path should indicate a strong functional connectivity, wherefore we consider the inverse of the graph weights for

its calculation. Its computation is carried out by the Dijkstra's algorithm (Rivest et al., 2000), which requires the weights to be positive.

Based on the shortest paths of a network we calculate closeness centrality $C_w(v)$ - a measure that indicates how often a node $v$ participates in all shortest paths of the graph. It is given by:

$$C_w(v) = \frac{n-1}{\sum_{u \in V \setminus \{v\}} dist_w(u,v)} \tag{3.4}$$

Here, $n$ is the number of all nodes $V$ in graph.

**Clustering coefficient**

The clustering coefficient $cc(v)$ describes to what degree the neighbors of a node $v$ are connected among each other and with node $v$. Since our network is weighted, we use the Zhang-Horvath clustering coefficient (Zhang & Horvath, 2005; Kalna & Higham, 2007), which is an extension to the 'standard' algorithm applied to binary graphs:

$$cc(v) = \frac{\sum_{i \neq v} \sum_{j \neq v, j \neq i} \hat{w}_{vi} \hat{w}_{ij} \hat{w}_{vj}}{\Big( \sum_{j \neq v} \hat{w}_{vj} \Big) \Big( \sum_{j \neq v} \hat{w}_{vj} - 1 \Big)} \tag{3.5}$$

for $i$, $j$ neighbors of $v$ and $\hat{w}$ denoting the weights normalized by the highest weight in the network, such that $0 \leq \hat{w} \leq 1$.

**Modularity**

A graph can be partitioned into smaller components. Modularity measures the deviation of the properties of these components as compared to the components of a random graph with the same edge weight. Accordingly, the modularity of a partition $p$ of a network $G$ into communities $c$ is given by (Newman, 2004)

$$Q(p) = \frac{1}{2m} \sum_{i,j \in V} \big( w_{ij} - \frac{deg_w(i) \cdot deg_w(j)}{2m} \big) \delta_{c_i c_j} \tag{3.6}$$

where $\delta_{c_i c_j}$ is 1 , if community $c_i$ of node $i$ is the same as community $c_j$ of node $j$ and 0 otherwise and $m = \frac{1}{2} \sum_{i,j \in V} w_{ij}$ is the total sum of edge weights in a network. Although there are many different definitions in literature about what a community comprises of , we define a community as a group of strongly interconnected nodes that make only weak connections to other communities. In addition a node can maximally contribute to one community, hence we want to find the partition that maximizes modularity. Because this is computationally very demanding, it is important to use a very effective algorithm. We therefore use the fast algorithm by (Blondel et al., 2008), which is implemented in the Python packages community.

Unfortunately this implementation is only suited for undirected graphs, so we investigate modularity only for these type of graphs.

### 3.4.6 Statistical model

The graph data generated with a graph-construction method are used within an exchangeable parametric statistical model. Let us recall that the purpose of the fMRI scan of a patient is to give the clinician a likelihood for the patient's health condition,

$$\mathrm{P}(\text{graph data from fMRI scan} \,|\, \text{health condition} \,\wedge\, \text{prior info}), \tag{3.7}$$

which she combines with the likelihoods from other tests and her initial probability assignment, to obtain via Bayes's theorem a final probability for the health condition (Sox et al., 2013):

$$\overbrace{\mathrm{P}(\text{health condition} \,|\, \text{results of all tests} \,\wedge\, \text{prior info})}^{\textit{final probability}} \propto$$
$$\textit{likelihoods} \begin{cases} \mathrm{P}(\text{graph data from fMRI scan} \,|\, \text{health condition} \,\wedge\, \text{prior info}) \\ \times\, \mathrm{P}(\text{results of other tests} \,|\, \text{health condition} \,\wedge\, \text{prior info}) \\ \times\, \cdots \end{cases}$$
$$\times\, \underbrace{\mathrm{P}(\text{health condition} \,|\, \text{prior info})}_{\textit{initial probability}}. \tag{3.8}$$

The prior information also includes test results from previous patients, so that the prediction becomes the more accurate and reliable the more patients have been previously observed.

The functional dependence of the likelihood on the graph data is determined by the statistical model we use, and may be different for each health condition. The statistical model is determined by additional assumptions or hypotheses. Such hypotheses and the functional form of the likelihood may depend on the particular space of graph data (e.g., real-valued, or positive, or bounded within a finite range, or combinations thereof), and therefore on the graph-construction method.

As explained in Sec. 3.2.3, our purpose is to assess as much as possible the relative predictive power of the different graph-construction methods. We therefore would like the functional dependence on the graph-data space to be minimal. In the present study we adopt the working hypothesis that only the first and second empirical moments – means and correlations – of the graph data from past patients with the same health condition are relevant to make predictions about a new patient. This hypothesis is adopted for all graph-construction methods. We also assume our initial knowledge of the graph data to be approximately invariant under rescalings of their values (Minka, 2001). Finally, we do not take into account the natural range of variability (positive, bounded, etc.) of the graph data; this choice does not seem to impact the predictive power of the model (Porta Mana et al., 2018).

These assumptions almost uniquely determine the statistical model and the likelihood (Porta Mana et al., 2018): it turns out to be a multivariate t distribution (Kotz & Nadarajah, 2004; Minka, 2001; Murphy, 2007). More precisely: Select a particular health condition, e.g. Alzheimer's disease. Denote with $\boldsymbol{f}_0$ the $d$-dimensional vector of graph data obtained from the patient's fMRI scan via a particular graph-construction method, and with $(\boldsymbol{f}_i)$ the graph data of $n$ previous patients with the selected health condition. Then the likelihood that the present patient has the selected health condition is

$$\mathrm{p}[\boldsymbol{f}_0 |\, (\boldsymbol{f}_i), \kappa_0, \boldsymbol{\delta}_0, \nu_0, \boldsymbol{\Delta}_0, M] \equiv \mathrm{p}(\boldsymbol{f}_0 |\, \kappa, \boldsymbol{\delta}, \nu, \boldsymbol{\Delta}, M) = \mathrm{t}\Big[\boldsymbol{f}_0 \Big|\ \nu - d + 1, \boldsymbol{\delta}, \tfrac{\kappa+1}{\kappa\,(\nu-d+1)}\boldsymbol{\Delta}\Big] \tag{3.9}$$

$$\text{with} \qquad \begin{aligned} &\kappa = \kappa_0 + n, && \nu = \nu_0 + n, \\ &\boldsymbol{\delta} = \frac{\kappa_0\,\boldsymbol{\delta}_0 + n\,\overline{\boldsymbol{f}}}{\kappa_0 + n}, && \boldsymbol{\Delta} = \boldsymbol{\Delta}_0 + n\ \mathrm{Cov}(\boldsymbol{f}) + \frac{\kappa_0\, n}{\kappa_0 + n}\big(\overline{\boldsymbol{f}} - \boldsymbol{\delta}_0\big)\big(\overline{\boldsymbol{f}} - \boldsymbol{\delta}_0\big)^{\intercal}, \end{aligned} \tag{3.10}$$

where t is a multivariate t distribution with $\nu - d + 1$ degrees of freedom, mean $\boldsymbol{\delta}$, and scale matrix $\frac{\kappa+1}{\kappa\,(\nu-d+1)}\boldsymbol{\Delta}$, and

$$\overline{\boldsymbol{f}} \coloneqq \frac{1}{n}\sum_i \boldsymbol{f}_i, \qquad \mathrm{Cov}(\boldsymbol{f}) \coloneqq \frac{1}{n}\sum_i (\boldsymbol{f}_i - \overline{\boldsymbol{f}})(\boldsymbol{f}_i - \overline{\boldsymbol{f}})^{\intercal} \tag{3.11}$$

are the empirical mean and covariance matrix of the previous graph data.

The parameters $\kappa_0$, $\nu_0$, $\boldsymbol{\delta}_0$, $\boldsymbol{\Delta}_0$ should reflect our initial knowledge of the graph parameters. For the reasons explained above and in Sec. 3.2.3, we fix one set of values identical for all graph-construction methods: $\kappa = 1$, $(\boldsymbol{\delta}_0)_a = 0.5$, $\boldsymbol{\Delta}_0 = 2.5\boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix. These values yield an initial distribution (before any data from previous patients) centered on positive values of unit order of magnitude, as all the graph data indeed are for each graph-construction methodionship .

### 3.4.7 Supportive evaluation measures of graph construction methods

#### 3.4.7.1 Significance test

We measure the significance level of the mean values of a graph property distribution between pairs of the three healthy conditions (control-AD, control-MCI, MCI-AD) based on the Student's t-test, if variances are equal (F-test ), and Welch's t-test otherwise. The underlying null hypothesis is that the means of the two data arrays are assumed to be equal, which is rejected for p-values smaller then 0.05.

#### 3.4.7.2 Dendrograms of subject order

Subjects indexed from 1 to 56 (total number of participants) across all healthy conditions are ordered according to the mean values of the certain graph properties distribution. The indices of the ordering (the rank) calculated for each graph construction method is then used

in order to construct the dendrogram. In the dendrogram the Euclidean distance between two indices arrays is indicated by the height of the top of the U-link linking the two arrays. In addition indices arrays with a small distance are clustered together.

#### 3.4.7.3 Support vector machines

For all complete graphs constructed by all different graph construction methods we apply a support vector classification (Python: *sklearn.svm.SVC*) on each pair of health conditions (control-AD, control-MCI, MCI-AD). Hereby we choose the graph properties such that the performance of the algorithm maximizes. We use the default parameters and do not optimize performance by varying the kernel coefficient or the penalty parameter of the error term.

## 3.5 Supplementary Tables and Figures

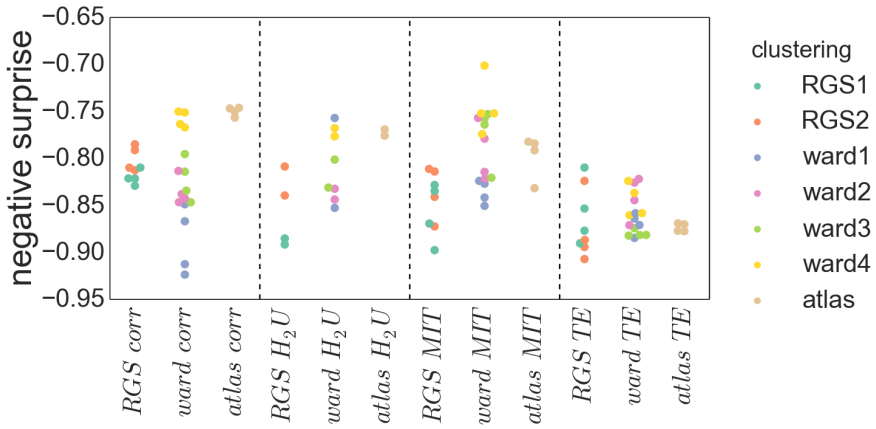


**Figure 3.14: Negative surprise of the different graph construction methods.** Figure identical with fig. 3.8 but with different clustering submethods depicted in different colors (see legend) .

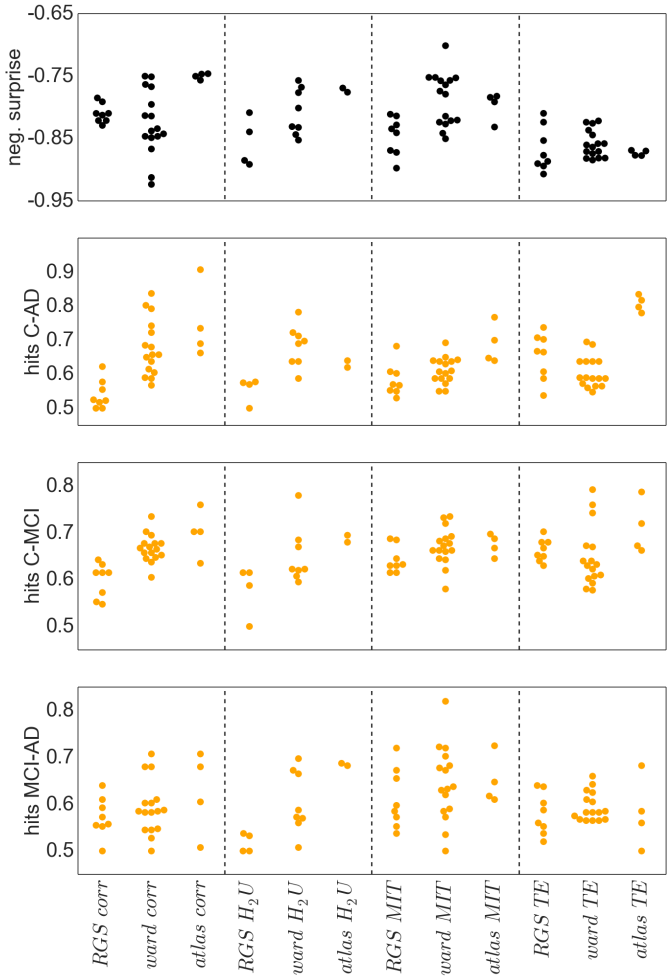


**Figure 3.15: Negative surprise and classification results based on support vector machines** First panel: Negative surprise as depicted in fig. 3.15. Note: it is calculated based on all three health conditions. Other panels: Correct hits based on support vector machine classification for control-AD (second panel), control-MCI (third panel) and MCI-AD (last panel).

# Chapter 4

# Firing rate homeostasis counteracts changes in stability of recurrent neural networks caused by synapse loss in Alzheimer's disease

The impairment of cognitive function in Alzheimer's disease is clearly correlated to synapse loss. However, the mechanisms underlying this correlation are only poorly understood. Here, we investigate how the loss of excitatory synapses in sparsely connected random networks of spiking excitatory and inhibitory neurons alters their dynamical characteristics. Beyond the effects on the network's activity statistics, we find that the loss of excitatory synapses on excitatory neurons shifts the network dynamic towards the stable regime. The decrease in sensitivity to small perturbations to time varying input can be considered as an indication of a reduction of computational capacity. A full recovery of the network performance can be achieved by firing rate homeostasis, here implemented by an up-scaling of the remaining excitatory-excitatory synapses. By analysing the stability of the linearized network dynamics, we explain how homeostasis can simultaneously maintain the network's firing rate and sensitivity to small perturbations.

## 4.1 Introduction

Accelerated synapse loss is a prominent feature in many types of neurodegenerative disorders, such as Huntington's disease, frontotemporal dementia or Alzheimer's disease (Zhan et al., 1993; Brun et al., 1995; Morton et al., 2001; Lin & Faber, 2002; Scheff et al., 2014). In Alzheimer's disease (AD), synapse loss appears to be particularly important, as it is widespread across different brain areas and constitutes a key marker in the AD pathology (see e.g. Scheff et al., 2014).

The mechanisms underlying AD related synaptic modifications are currently the subject of intensive research, which has revealed that a number of different alterations at the molecular level may ultimately lead to synaptic decay (Sheng et al., 2012; Dorostkar et al., 2015; Tampellini, 2015), such as an abnormal occurrence of oligomeric and aggregated $\beta$-amyloid-peptides (A$\beta$), an abnormal phosphorylation of the tau protein and the occurence of neurofibrillary tangles, and a disrupted signaling in neuroinflammatory and oxidative stress responses (Tampellini, 2015; Tönnies & Trushina, 2017; Frere & Slutsky, 2017; Rajendran & Paolicelli, 2018).

Previous studies have uncovered a strong positive correlation between cognitive impairment in AD patients and synapse loss (DeKosky & Scheff, 1990; Scheff et al., 1990; Terry et al., 1991; Scheff & Price, 1993; Masliah et al., 1994; Scheff & Price, 2003; Scheff & Douglas, 2006; Scheff et al., 2011). In contrast, correlations between the cognitive status and the density of plaques or tangles have frequently been reported as rather weak. Synapse loss is therefore not merely a structural epiphenomenon of AD, but appears to be *the* physical correlate of cognitive decline.

While the most commonly reported early symptom of AD is memory deterioration, the disease is associated with a wide range of other cognitive problems such as stereotyped, repetitive linguistic production, visuo-spatial deficits and disorientation, apraxia, and loss of executive functions, i.e. planning and abstract reasoning (Bennett et al., 2002; Weintraub et al., 2012). The observed progression of cognitive symptoms goes hand in hand with brain tissue atrophy (Smith, 2002a; de Toledo-Morrell et al., 2000; Thompson et al., 2003) associated with loss of synapses (Chen et al., 2018), suggesting that the synaptic degeneration may underlie the cognitive deterioration following the gradual involvement of different, functionally specialized brain regions.

However, mechanisms exist that counteract synapse loss (Small, 2004; Fernandes & Luísa, 2016), at least in the early stages of the disease. Various studies have shown that the loss of synapses is accompanied by a growth of remaining synapses, such that the total synaptic contact area (TSCA) per unit volume of brain tissue is approximately preserved (DeKosky & Scheff, 1990; Scheff & Price, 2003; Scheff & Douglas, 2006; Neuman et al., 2014). It is likely that such compensatory mechanisms underlie the observed delay in the onset of cognitive symptoms with respect to the onset of symptoms at the cellular level (Morris, 2005).

The heterogeneity in the disease progression and the propensity to transition from healthy cognitive aging to mild cognitive impairment and dementia may thus be associated to a subject's ability to counteract synapse loss and, to a certain extent, maintain global functionality in a way that masks the progressive underlying pathophysiology. Such homeostatic, regulatory mechanisms appear to play an important role in counteracting structural deterioration and preserving computational capabilities. On the other hand, they pose important challenges to the network's functionality since they have the potential to disrupt the specificities of a circuit's microconnectivity (namely the distribution of synaptic strengths) and thus degrade its information content (e.g. Fröhlich et al. 2008). Successful homeostatic compensation thus requires a balanced orchestration which preserves the system's computational properties and macroscopic dynamics, e.g., average firing rates (Lütcke et al., 2013; Slomowitz et al., 2015)

and E/I balance (Zhou & Yu, 2018), as well as the relative ratios and distributions of synaptic strengths (e.g. synaptic scaling mechanisms; Keck et al. 2013; Vitureira & Goda 2013).

Understanding the circuit-level consequences of synaptic alterations, entailing both the deregulation by synapse loss and recovery through homeostasis, is essential to understand whether they represent a negative symptom of the disease or a compensatory response. One likely effect is the modification of the network's firing rate. In order to maintain a physiological operating regime far from activity extremes (quiescence or epileptic activity), a network needs the capacity to regulate its firing rate.

The degree to which this may be impaired in AD is still under debate (see Styr & Slutsky, 2018; Frere & Slutsky, 2017). Whereas the effects of synaptic alterations on the network dynamics have been partially characterized, a direct link between synapse loss, network dynamics and functional decline has yet to be systematically established, with only a few studies addressing the topic (Horn et al., 1993, 1996; Ruppin & Reggia, 1994). However, this connection may prove fruitful, both for understanding the disease itself and for fostering the development of new diagnostic and therapeutic approaches.

It is currently unknown to what extent homeostatic mechanisms, such as increasing the synaptic area (DeKosky & Scheff, 1990; Scheff & Douglas, 2006), can completely recover the neuronal network's firing rate, nor whether the preservation of the firing rate by such mechanisms entails the preservation of cognitive performance. In this study, we investigate the link between structure, dynamics and function using a recurrent spiking neural network model (Brunel & Hakim, 1999).

Despite their simplicity, such systems have been shown to support computations, such as e.g. stimulus categorization, associative learning and memory, information routing and propagation, etc. (see, e.g. Jaeger & Haas, 2004; Eliasmith & Anderson, 2004; Maass et al., 2002; Buesing et al., 2011; Boerlin et al., 2013; Abbott et al., 2016). Additionally, although these models have complex behavioral repertoires, they are often simple enough that their dynamics can be assessed analytically. The stability of the dynamics can then be related to computational task performance, such as the network's sensitivity to perturbation (Legenstein & Maass, 2007a,b). Thus, an analytical treatment of network dynamics can provide insight into why some realizations of such networks perform better than others and how performance is affected by structural changes. Theoretical studies explicitly addressing this issue have so far focused either on the disruption of oscillations or functional connectivity of the whole brain, or on memory only (especially memory retrieval; Horn et al., 1993, 1996; Ruppin & Reggia, 1994).

Here, we investigate how the loss of excitatory-excitatory synapses in sparsely connected random networks of spiking excitatory and inhibitory neurons (Sec. 4.2.1) and firing rate homeostasis, based on upscaling the remaining excitatory-excitatory connections, alters the dynamical characteristics of a network. Surprisingly, we find that firing rate homeostasis can restore a variety of dynamical features caused by synaptic loss, including the increase in spike train regularity, the drop in the fluctuations of population activity and the reduction of the synaptic contact area (Sec. 4.2.2) caused by synaptic loss. In addition, we observe that synaptic loss decreases the network's sensitivity to small perturbations (Sec. 4.2.3), such

that a network operating near the 'edge of chaos' would be shifted by synaptic loss to a more stable regime; a shift which has been shown in previous studies to result in a decrease in computational capacity (Langton, 1990; Legenstein & Maass, 2007b,a; Schrauwen et al., 2009; Dambre et al., 2012; Schuecker et al., 2017), and may account for the cognitive deficits observed in Alzheimer's disease. Here, too, firing rate homeostasis counteracts the shift towards the stable regime. We further show that these compensatory mechanisms ultimately become exhausted if physiological limits are placed on the growth of the synapse. As it is not obvious why simply maintaining the firing rate also maintains the stability of the network, we analyze the stability of the linearized network dynamics and discover a strictly monotonic relationship between the firing rate and the spectral radius of the network, which explains the restoration of the dynamics under the influence of firing rate homeostasis (Sec. 4.2.3).

## 4.2 Results

### 4.2.1 Computational network model of Alzheimer's disease

We study the effects of AD related synaptic alterations on the network dynamics and computational characteristics in the framework of a generic mathematical neuronal network model (Fig. 4.1 A), which captures prominent structural and dynamical features of local neocortical networks such as the relative numbers of excitatory and inhibitory neurons (Scholl, 1956; Abeles, 1982) and synapses (DeFelipe & Fariñas, 1992; Gulyás et al., 1999), sparse connectivity (Abeles, 1982; Binzegger et al., 2004), small synaptic weights (Lefort et al., 2009), irregular (Tomko & Crapper, 1974; Softky & Koch, 1993; Shadlen & Newsome, 1998) and predominantly asynchronous spiking (Ecker et al., 2010), large membrane potential fluctuations (Petersen & Crochet, 2013; Cowan & Wilson, 1994; Timofeev et al., 2001; Steriade et al., 1993), and a tight dynamical balance between excitatory and inhibitory synaptic currents (Okun & Lampl, 2008).

The network is composed of randomly and sparsely connected populations of excitatory (E) and inhibitory (I) integrate-and-fire neurons, driven by external spiking input. The overall coupling strength is determined by the reference synaptic weight $J$. For simplicity, all excitatory connections (EE and IE) and all inhibitory connections (EI and II), respectively, have equal synaptic weight: $J_{\mathrm{EE}} = J_{\mathrm{IE}} = J$ and $J_{\mathrm{EI}} = J_{\mathrm{II}} = -gJ$ in the intact network (i.e. before synapse loss). The relative strength $g$ of inhibitory weights is chosen such that the network is dominated by inhibition, to permit asynchronous irregular firing at low rates (Brunel, 2000). A complete specification of the network model and parameters can be found in Sec. 4.4.1 and in the Supplementary Material (Sec. 4.5.1, Sec. 4.5.2). An illustration of the connectivity of the excitatory population for an intact network and an example spike train is given in Fig. 4.1 B.

We implement the effects of AD on the network connectivity by reducing the number of excitatory synapses on excitatory neurons (EE synapses; Lacor et al., 2007; Dorostkar et al., 2015), whilst keeping the number of connections between other populations (EI, IE, II)

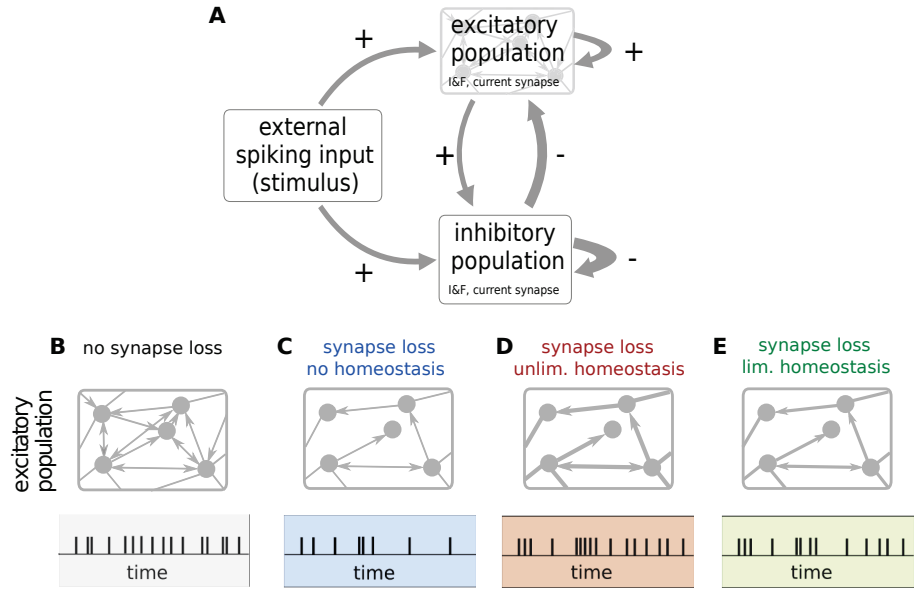


**Figure 4.1: Sketch of the network model of Alzheimer's disease and homeostasis. A)** The network comprises two reciprocally and recurrently connected populations of excitatory (E) and inhibitory (I) integrate-and-fire neurons, excited by an external spiking input. Thickness of arrow indicates relative strength of the connection. In this study, Alzheimer's disease is modeled by removing connections between excitatory neurons (loss of EE synapses) and upscaling of the remaining EE synapses to maintain the average firing rate (firing rate homeostasis). **B–E)** Sketch of EE connection density (number of arrows in upper panels), connection strength (thickness of arrows in upper panels) and resulting single-neuron spiking activity (lower panels). **B)** Intact network (without synapse loss). **C)** Synapse loss without homeostasis: removal of EE synapses and resulting reduction in firing rate. **D)** Synapse loss with unlimited homeostasis: removal of EE synapses and increase in strength of remaining EE synapses to maintain the average firing rate. Synaptic weights are allowed to grow without bounds. **E)** Synapse loss with limited homeostasis: removal of EE synapses and bounded increase in strength of remaining EE synapses. Here, synaptic weights cannot exceed $120\%$ of their reference weight. The firing rate is therefore only partially recovered. For a complete description and parameter specification of the network model, see Sec. 4.4.1 and Supplementary Material (Sec. 4.5.1, Sec. 4.5.2).

constant. In the absence of any compensation mechanism, this modification leads to a reduction in the average firing rate (see Sec. 4.2.2).

In biological neuronal networks, long-term activity levels are often stabilized by homeostatic regulation (Bennett et al., 2002; Turrigiano, 2008; Marder & Goaillard, 2006). While a maintenance of firing rates has been observed at the level of individual neurons (Lütcke et al., 2013), long-term recordings suggest a predominance of a network-wide regulation (Slomowitz et al., 2015) targeting a constant population firing rate. Such a homeostatic stabilization of the population firing rate can be accounted for by a global adjustment of synaptic weights (synaptic scaling; Vitureira et al., 2012; Turrigiano, 2012). Indeed, in the early stages of AD, synapse loss seems to be compensated by a growth of the remaining synapses (DeKosky & Scheff, 1990; Scheff & Douglas, 2006; Neuman et al., 2014). To realize this mechanism in our spiking neuronal network, we implement a firing-rate homeostasis which compensates for the loss of EE synapses by a global increase in the weights $J_{\mathrm{EE}}$ of the remaining EE synapses, thereby preserving the population firing rate.

For advanced AD, where a large portion of the EE synapses has been lost, a full recovery of the population firing rate through synaptic scaling would require unrealistically large synaptic weights. During aging and dementia, the maximum increase in synaptic size has been reported to be in the range from $9\%$ to $24\%$ (see Scheff & Price, 2003, and references therein). We incorporate these findings by introducing an optional upper bound for the weight $J_{\mathrm{EE}}$ of EE synapses.

To uncover the differential effects of excitatory synapse loss and homeostasis, in this study we investigate the dynamical and computational characteristics of a network for three different scenarios: synapse loss without homeostatic compensation (Fig. 4.1 C), synapse loss with an unlimited firing rate homeostasis where synaptic weights can grow without bounds (Fig. 4.1 D), and synapse loss with limited firing rate homeostasis where the synaptic weights cannot exceed $120\%$ of the weight in the intact reference network (Fig. 4.1 E).

Note that the model's high level of abstraction enables us to identify fundamental mechanisms, to reduce the risk of overfitting, and to arrive at general conclusions that may be transferred to other brain regions or even different spatial scales. Empirically observed features of biological neural networks such as heavy-tail synaptic weight distributions (Song et al., 2005; Ikegaya et al., 2013) or active dendritic processing (Major et al., 2013) are not explicitly incorporated. As a consequence, model parameters such as synaptic weights have to be regarded es "effective" parameters and cannot be mapped to biological parameters in a one-to-one fashion. Selecting a particular set of parameters to be considered "biologically realistic" would be misleading. Therefore, rather than focusing on a specific configuration of the model, we systematically vary both the reference synaptic weight $J$ and the extent of synapse loss to uncover the general relationship between these parameters and the dynamical and computational properties of the network.

### 4.2.2 Total synaptic contact area and firing statistics

In the absence of homeostatic compensation (left column of Fig. 4.2), removal of excitatory synapses on excitatory neurons naturally results in a decrease in the population firing rate $\nu$, irrespective of the synaptic-weight scale $J$ (Fig. 4.2A). An upscaling of the remaining EE synapses (middle column) allows us to preserve the population firing rate, even if substantial amounts of synapses are removed (vertical contours in Fig. 4.2B). If the maximum synaptic weight is limited, firing rates are preserved only up to a critical level of synapse loss (early stages of AD; Fig. 4.2C).

Experimental studies have shown that, in early AD, the reduction in the number of synapses is accompanied by a growth of the remaining synapses such that the total synaptic contact area (TSCA) per unit volume is approximately preserved (DeKosky & Scheff, 1990; Scheff & Douglas, 2006; Neuman et al., 2014). Our simple AD network model reproduces this finding if we define the TSCA as the product of the number of EE connections and the synaptic weight $J_{\mathrm{EE}}$ (Sec. 4.4.2). Without homeostatic upscaling of EE weights, the TSCA is proportional to the number of EE connections and therefore quickly decreases with increasing levels of synapse loss (Fig. 4.2G). In the presence of firing rate homeostasis, however, the TSCA remains largely constant unless a majority of synapses is lost (Fig. 4.2H) or the maximum synaptic weight is reached (Fig. 4.2I). We conclude that the experimentally observed stabilization of the TSCA in the face of synapse loss may be a consequence of a homeostatic synaptic scaling regulated by the average population firing rate.

In physiologically relevant low activity regimes, neuronal firing is determined both by the mean as well as by fluctuations in the synaptic input. A reduction in the number of synapses followed by an upscaling of synaptic weights may preserve the average population firing rate; it cannot, however, simultaneously preserve the mean and the variance of the synaptic input currents. The neurons' working point, i.e. the statistics of the synaptic input, will inevitably change. A priori, it is therefore not clear to what extent synapse loss and firing rate homeostasis alter the overall firing statistics in the recurrent network beyond the average firing rate. Here, we address this question by studying the irregularity of spike generation by individual neurons, measured by the coefficient of variation CV of the inter-spike interval distribution, and spike-train synchrony, assessed by the normalized variance of the population spike count, the Fano factor FF, in $10\,\mathrm{ms}$ time intervals (see Sec. 4.4.2). Without homeostatic compensation, synapse loss generally results in spike patterns that are less irregular (Fig. 4.3A) and less synchronous (Fig. 4.3D). In the presence of firing rate homeostasis, however, both the CV and the FF are largely preserved (Fig.fig:statsplotsB,E). Only if the level of synapse loss becomes too severe or if the synaptic-strength limits are reached (limited homeostasis), the CV and the FF are reduced (Fig. 4.3B,C and Fig. 4.3E,F).

For illustration, Fig. 4.4 depicts the spiking activity for four example parameter settings marked by the symbols in Fig. 4.2 and Fig. 4.3. As Fig. 4.3 E and H already suggest, the overall spiking activity, e.g the number and the duration of synchronous event and the spiking frequency of single neurons, of the homeostatic network (Fig. 4.4 C) and the reference network (Fig. 4.4 A) are very similar. Only the exact timing of the synchronous events and the single neuron spiking differ. In the AD network without homeostasis (Fig. 4.4 B), the firing rates of both excitatory and inhibitory neurons are decreased. The number of synchronous

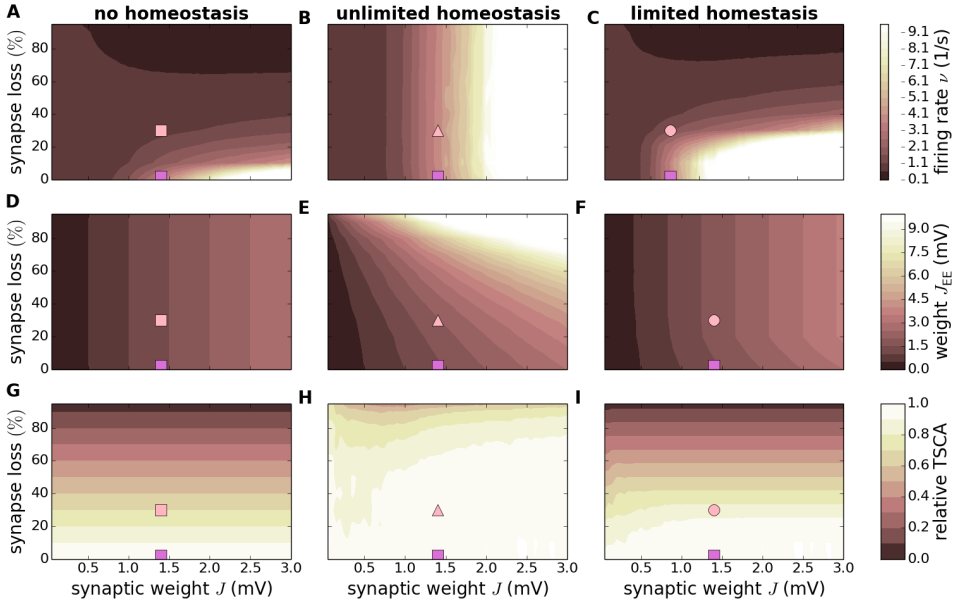


**Figure 4.2: Effect of synapse loss and firing rate homeostasis on firing rate, synaptic weights and total synaptic contact area.** Dependence of the time and population averaged firing rate $\nu$ (**A**–**C**), synaptic weight $J_{EE}$ (**D**–**F**) and the relative total synaptic contact area (TSCA) of EE synapses (**G**–**I**) on the reference weight $J$ and the degree of EE synapse loss in the absence of homeostatic compensation (left column), as well as with unlimited (middle column) and limited firing rate homeostasis (right column). Color-coded data represent mean across 10 random network realizations. Symbols mark parameter configurations shown in Fig. 4.4.

events, compared with the reference network (Fig. 4.4 A), does not seem to be decreased, but their duration does. The network with limited homeostasis (Fig. 4.4 D) is more similar to the AD network without homeostasis than the unlimited homeostasis networks, because the restriction in synaptic growth prevents the rate from being recovered.

### 4.2.3 Perturbation sensitivity and linear stability

An open, yet chronically ignored, question in Alzheimer's disease research is how cellular damage such as synapse loss affects patients' cognitive capabilities. A number of theoretical studies have shown that recurrent neuronal networks exhibit optimal computational performance characteristics for a variety of task modalities if they operate in a dynamical regime where small perturbations are neither instantly forgotten nor lead to entirely different network states (Langton, 1990; Legenstein & Maass, 2007b,a; Schrauwen et al., 2009; Dambre et al., 2012; Schuecker et al., 2017). In dynamical systems theory, this regime has been termed the "edge of chaos" as it represents the transition from a stable state with a low sensitivity to small perturbations to a chaotic state where the sensitivity to small perturbations is high.

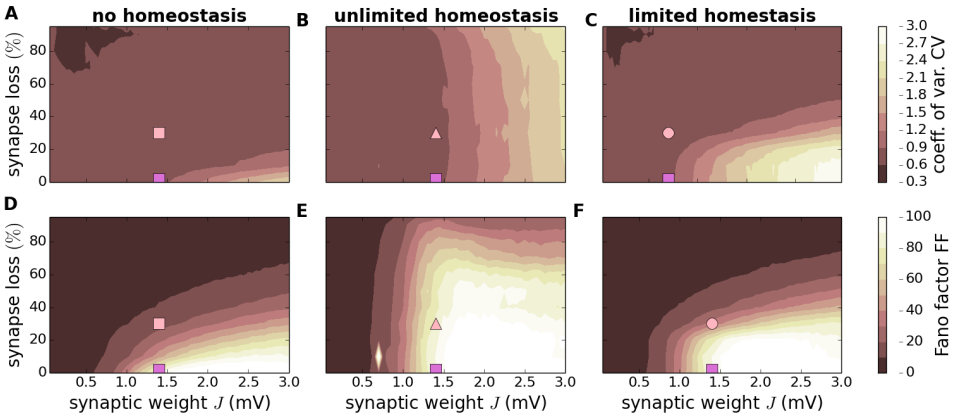


**Figure 4.3: Effect of synapse loss and firing rate homeostasis on spike train statistics.** Dependence of the coefficient of variation CV of inter-spike intervals (**A**–**C**) and the Fano factor FF of the population spike count (binsize $b = 10\,\mathrm{ms}$; **D**–**F**) on the synaptic reference weight $J$ and the degree of EE synapse loss in the absence of homeostatic compensation (left column), as well as with unlimited (middle column) and limited firing rate homeostasis (right column). Color-coded data represent mean across $10$ random network realizations. Symbols mark parameter configurations shown in Fig. 4.4.

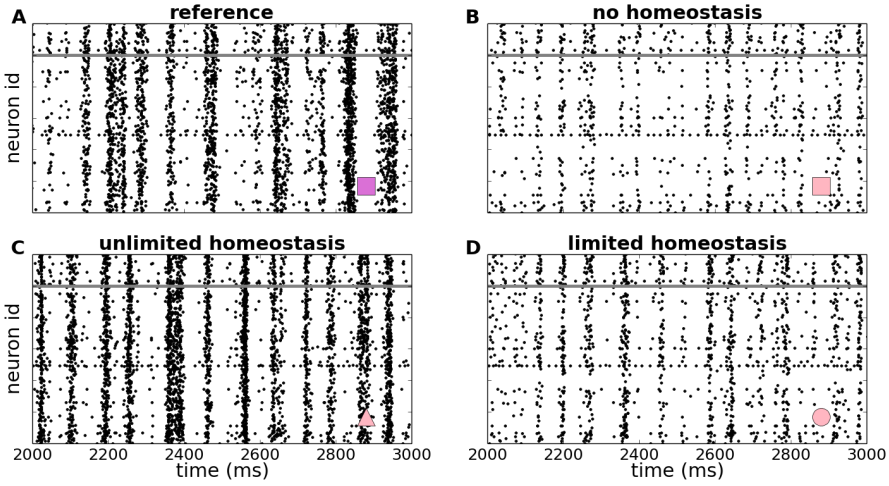


**Figure 4.4: Effect of synapse loss and firing rate homeostasis on spiking activity.** Spiking activity (dots mark time and sender of each spike) in an intact reference network (no synapse loss, $J_{\mathrm{EE}} = 1.4\,\mathrm{mV}$; **A**), as well as in networks where $30\%$ of the EE synapses are removed: **B)** no homeostasis ($J_{\mathrm{EE}} = 1.4\,\mathrm{mV}$), **C)** unlimited homeostasis ($J_{\mathrm{EE}} = 2.02\,\mathrm{mV}$), **D)** limited homeostasis ($J_{\mathrm{EE}} = 1.68\,\mathrm{mV}$). In all panels, the synaptic-weight scale is set to $J = 1.4\,\mathrm{mV}$. Examples depict parameter configurations marked by corresponding symbols in Fig. 4.2 and Fig. 4.3 (cf. marker in lower right corner of each panel). Regions below and above the gray horizontal line show spiking activity of a subset of $100$ excitatory and $25$ inhibitory neurons, respectively.

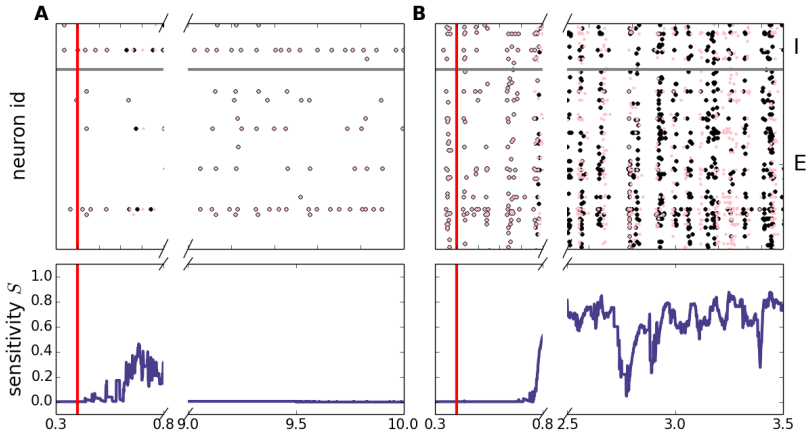


**Figure 4.5: Perturbation sensitivity.** Top: Example spiking activity (dots mark time and sender of each spike) of two identical networks (identical neuron parameters, connectivity, external input, initial conditions) with (black dots) and without perturbation (purple dots). The perturbation consists in delaying one external input spike at time $t^* = 400$ ms by $\delta t^* = 0.5$ ms. The vertical red line marks the time of the perturbation. Spikes of only $10\%$ of all neurons are shown. Neurons below and above the horizontal gray line correspond to excitatory and inhibitory neurons, respectively. Bottom row: Perturbation sensitivity $S(t) = 1 - |R(t)|$ obtained from the correlation coefficient $R(t)$ of the low-pass filtered spike trains generated by the unperturbed and the perturbed network (black and purple dots in top panels; see Methods Sec. 4.4.2). **A)** Stable dynamics ($J = 0.45$ mV, $K_{\mathsf{EE}} = 100$). **B)** Chaotic dynamics ($J = 1.75$ mV, $K_{\mathsf{EE}} = 100$). Note different time scales in A and B.

Here, we investigate the role of synapse loss and firing rate homeostasis for the network's sensitivity to perturbations as an indicator of its overall computational performance.

To assess the perturbation sensitivity, we simulate a given network twice with identical initial conditions and identical realizations of external inputs. In the second run, we apply a small perturbation by delaying one of the external input spikes to a single neuron by a fraction of a millisecond (Fig. 4.5). In stable regimes, the effect of this perturbation on the spiking response is transient and quickly vanishes (Fig. 4.5 A, top). In chaotic regimes, in contrast, the small perturbation leads to diverging spike patterns (Fig. 4.5 B, top). We quantify the network's perturbation sensitivity $S = 1 - |R|$ in terms of the long-term correlation coefficient $R$ between the low-pass filtered spike responses in the two runs (Fig. 4.5, bottom). With this definition, $S = 0$ and $S = 1$ correspond to insensitive (stable) and highly sensitive (chaotic) networks, respectively (for details, see Sec. 4.4.2).

For small synaptic weights $J$, the network dynamics is always stable ($S = 0$) for our choice of parameters, irrespective of the degree of synapse loss and the absence or presence of homeostatic compensation (Fig. 4.6). In this regime, the perturbation has no long-term effect: after a transient phase, the response spike patterns in the perturbed and the unperturbed simulation are exactly identical (at the temporal resolution $\Delta t_{\mathrm{f}} = 1$ ms of the recorded signals).

The intact networks (zero synapse loss) enter a chaotic regime ($S > 0$) if the synaptic weights $J$ exceed a certain critical value. Removal of EE-synapses without homeostatic compensation leads to a shift of this transition towards larger synaptic weights (Fig. 4.6 A). Networks in the chaotic regime eventually become insensitive to perturbations with progressing EE-synapse loss. In the presence of firing rate homeostasis, in contrast, the perturbation sensitivity is preserved (color gradient in Fig. 4.6 B is predominantly left to right, rather than top to bottom). Unless the homeostatic strengthening of EE-synapses is limited (limited homeostasis; Fig. 4.6 C), this maintenance of the perturbation sensitivity is observed even if the degree of synapse loss is substantial ($> 80\%$).

We conclude that synapse loss, as observed in Alzheimer's disease, tends to reduce the perturbation sensitivity of the affected networks, and may thereby impair their computational performance for a broad range of task modalities. Homeostatic mechanisms that preserve the average network activity (firing rate) can prevent this reduction in sensitivity and, hence, the decline in computational capability.

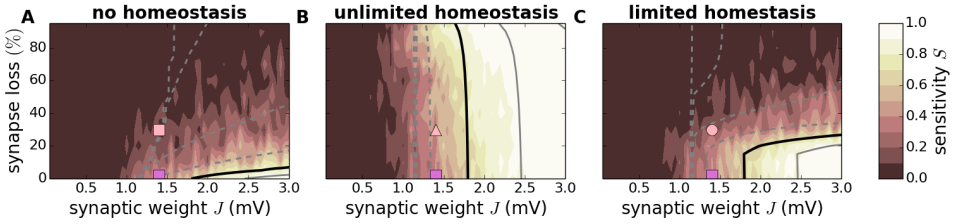


**Figure 4.6: Effect of synapse loss and firing rate homeostasis on perturbation sensitivity.** Dependence of perturbation sensitivity $S$ on the synaptic reference weight $J$ and the degree of EE synapse loss in the absence of homeostatic compensation (**A**), as well as for unlimited (**B**) and limited firing rate homeostasis (**C**). Color-coded data represent mean across $10$ random network realizations. Superimposed black and gray curves mark regions where the linearized network dynamics is stable (gray dashed; spectral radius $\rho = \ldots, 0.6, 0.8$), about to become unstable (black; $\rho = 1$), and unstable (gray solid; $\rho = 1.2, 1.4, \ldots$). Pink symbols mark parameter configurations shown in Fig. 4.2 and Fig. 4.4.

So far, the reported results on the perturbation sensitivity were obtained by network simulations for a specific set of parameters. In the following, we employ an analytical approach; firstly, to show that our findings are general and do not depend on the details of the network model, and secondly, to shed light on the mechanisms underlying the reduction in perturbation sensitivity by synapse loss and its maintenance by firing rate homeostasis.

As shown in Sompolinsky et al. (1988), the dynamics of large random networks of analog nonlinear neurons without (or with constant) external input undergoes a transition from a stable to a chaotic regime at some critical synaptic coupling strength. The study further revealed that this transition coincides with a critical point where the local linearized network dynamics becomes unstable. For more realistic networks of spiking neurons, networks with fluctuating external input or networks with a more realistic connectivity structure, a strict correspondence between the onset of chaotic dynamics and linear instability could not be established (Ostojic, 2014; Engelken et al., 2015; Ostojic, 2015; Kadmon & Sompolinsky, 2015; Harish & Hansel, 2015; Schuecker et al., 2017). Nevertheless, various previous studies suggest

that the two transition types are interrelated, in the sense that a change in the linear stability characteristics is accompanied by a change in the network's sensitivity to small perturbations.

Here, we propose that the linear stability characteristics can serve as an indirect and easily accessible indicator of the network's sensitivity to small perturbations, and hence its computational capability. As described in Sec. 4.4.3, the linearized network dynamics is determined by the effective connectivity matrix $\boldsymbol{W}$. Its components $w_{ij} = \boldsymbol{W}_{ij}$ $(i, j \in \{1, \ldots, N\})$ measure the effect of a small fluctuation in the firing rate $\nu_j(t)$ of a presynaptic neuron $j$ on the rate $\nu_i(t)$ of the postsynaptic neuron $i$ at a specific working point determined by the stationary firing rates $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_N)$. The effective connection weights are hence determined not only by the synaptic weights $J_{ij}$, but also by the excitability of the target cell $i$, which is in turn determined by the statistics of the synaptic input fluctuations, i.e. the dynamical state of the local network. The linearized dynamics becomes unstable if the spectral radius $\rho = \mathrm{Re}(\lambda_{\max})$, the real part of the maximal eigenvalue $\lambda_{\max}$ of $\boldsymbol{W}$, exceeds unity.

Loss of EE synapses corresponds to setting a fraction of the excitatory components $w_{ij}$ $(i, j \in \mathcal{E})$ to zero. In the absence of homeostatic compensation, we expect this weakening of positive feedback to have a stabilizing effect. The dependence of the effective weights $w_{ij}$ on the working point, however, leads to a non-trivial effect of synapse loss and firing rate homeostasis on the spectral radius $\rho$. Here, we compute $\rho$ by employing the diffusion approximation of the leaky integrate-and-fire neuron and random-matrix theory (for details, see Sec. 4.4.3).

As shown in Fig. 4.6 (black and gray curves), the linear stability characteristics (as measured by the spectral radius $\rho$) bear striking similarities to the sensitivity to perturbations. In the absence of homeostasis, loss of EE synapses leads to a fast decrease in $\rho$. Linearly unstable networks quickly become stable (Fig. 4.6 A). Firing rate homeostasis, in contrast, preserves the spectral radius $\rho$, even if a substantial fraction of EE synapses is removed. Linearly unstable networks remain unstable (Fig. 4.6 B), until the homeostatic resources are exhausted (Fig. 4.6 C).

The analytical approach described in Sec. 4.4.3 provides us with an intuitive understanding of why and under what conditions firing rate homeostasis preserves the linear stability characteristics in the face of synapse loss. The analysis shows that, in the presence of firing-rate homeostasis, the spectral radius $\rho$ is uniquely determined by the stationary average firing rate (red curves and symbols in Fig. 4.7B and eq.4.17). For the parameters chosen in this study, an approximately unique dependence on the firing rate is also observed in the absence of homeostasis and for limited homeostasis (blue and yellow curves in Fig. 4.7B). Network simulations reveal similar findings for the perturbation sensitivity $S$ (Fig. 4.7A). For unlimited homeostasis, the firing rate, the perturbation sensitivity and the spectral radius remain (approximately) constant during synapse loss (red curves in Fig. 4.7). In the absence of homeostasis or for limited homeostasis, firing rates change; the corresponding spectral radii $\rho$ and perturbation sensitivities $S$ nevertheless remain within a narrow band (bue and yellow curves in Fig. 4.7). The number $K_{\mathrm{EE}}$ and the strength $J_{\mathrm{EE}}$ of EE synapses therefore play only an indirect role by determining the stationary firing rate $\nu$. Any combination of $K_{\mathrm{EE}}$ and $J_{\mathrm{EE}}$ that preserves $\nu$ will simultaneously preserve $\rho$ (and $S$).

The unique dependence of the spectral radius $\rho$ on the firing rate $\nu$ is a consequence of the working-point dependence of the effective weights $w_{ij} \approx \eta(\nu_i) J_{ij}/\sigma_i$, where $\eta(\nu_i)$ is a function of the firing rate $\nu_i$ of the target neuron $i$ (see eq. 4.15). To maintain the stationary firing rate $\nu_\mathrm{E}$ of excitatory neurons, the synaptic weights $J_\mathrm{EE}$ are increased to compensate for the loss of excitatory synapses, i.e. for the decrease in the number $K_\mathrm{EE}$ of excitatory inputs. This increase in the synaptic weights $J_{ij}$ (for neurons $i, j$ both in the excitatory population) is accompanied by an increase in the variance $\sigma_i^2$ of the synaptic input received by the target neuron $i$. If the response firing rate $\nu_i$ is kept constant (as is the case in the presence of firing rate homeostasis), an increase in $\sigma_i$ leads to a decrease in neuron $i$'s sensitivity to a modulation of the input current caused by a spike of the source neuron $j$. This interplay between an upscaling of the weights $J_{ij}$ and a downscaling of the neuron's modulation sensitivity restricts the growth in the effective weight $w_{ij}$, and, ultimately, leads to a preservation of the spectral radius $\rho$. In Sec. 4.4.3, we demonstrate this effect for a homogeneous network of leaky-integrate-and-fire neurons. The derivation relies on the assumption that the synaptic weights are sufficiently small and the rate of synaptic events is high (diffusion approximation), that the stationary firing rates $\nu_\mathrm{E}$ and $\nu_\mathrm{I}$ of excitatory and inhibitory neurons are identical (homogeneity), and that the input fluctuations caused by external sources are small compared to those generated by the local network.

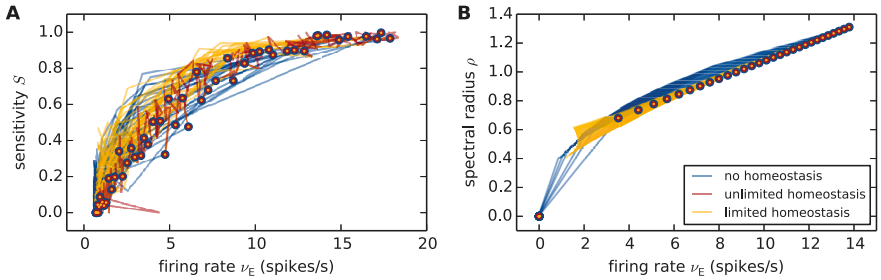


**Figure 4.7: Firing rate as predictor of perturbation sensitivity and linear stability.** Dependence of the perturbation sensitivity $S$ (**A**; simulation results) and the linear stability quantified by the spectral radius $\rho$ (**B**; theory) on the mean stationary firing rate $\nu_\mathrm{E}$ of the excitatory neuron population in the absence of homeostasis (blue), as well as for unlimited (red) and limited firing rate homeostasis (yellow). Each curve depicts data for a fixed reference weight $J \in \{0, \ldots, 3\}$ mV and various degrees of synapse loss from $0\%$ to $50\%$ (vertical paths in Fig. 4.6 and Fig. 4.8). For unlimited homeostasis, the firing rate and the spectral radius are nearly perfectly conserved while removing synapses. In B), the red curves are therefore too short to be visible. Circles depict results for $0\%$ synapse loss. Same data as in Fig. 4.6 and Fig. 4.8 D–F and M–O.

## 4.3 Discussion

In this article, we study the effect of Alzheimer's disease on the dynamics and perturbation sensitivity of recurrent neuronal networks. To this end, we employ a computational model of a generic neuronal network composed of excitatory and inhibitory spiking neurons. Alzheimer's disease is implemented in the form of a loss of excitatory synapses on to

excitatory neurons. The resulting decrease in the firing rate is avoided (or retarded) by firing rate homeostasis, which is achieved by increasing the weights of the remaining excitatory-excitatory (EE) synapses. In one scenario, we allow synaptic weights to grow without bounds; in another, to ensure that they stay within the physiological range (Scheff & Douglas, 2006), we limit the maximum synaptic weight during homeostasis to 120% of the reference weight in the intact network (i.e. before synapse loss).

We show that, in the absence of homeostatic compensation, a progressive loss of EE synapses not only reduces the average firing rate, but also leads to an increase in spike train regularity and a decrease in the fluctuations of the population activity. At first glance, the reduction in firing rate appears to be in contrast with the observation that the network activity of early affected areas (e.g. hippocampus) is enhanced (Mendez et al., 1994; Amatniek et al., 2006; Lam et al., 2017). However, this hyperactivity is mostly reported in very early preclinical disease stages, in which increasing oligomeric Amyloid-beta (A$\beta$) accumulates (Hall et al., 2015). A$\beta$ oligomers seem to enhance the occurence of phosphorylated tau in spines (for review see Tampellini, 2015), causing a degradation of excitatory-excitatory connections (Merino-Serrais et al., 2013). This might be the reason why during later disease stages, in which the tau-pathology becomes more prominent, the network activity decreases as predicted by our model (see, e.g., Dickerson et al., 2005; O'Brien et al., 2010; Herholz, 2010; Busche et al., 2012).

According to our AD model, the decrease in firing rate in more advanced disease stages can be delayed by homeostatic synaptic scaling. Moreover, our model predicts that as long as the homeostatic mechanisms are able to restore the network's firing rate, the CV and Fano factor are also preserved. Once these mechanisms are exhausted in the later disease stages, our model predicts that the spike train regularity increases and the fluctuations in the population activity decrease. That a weakening of synaptic coupling decreases the CV has also been found in other computational studies (Ostojic, 2014; Kriener et al., 2014); an experimental investigation of the evolution of activity statistics in AD animal models has, to our knowledge, yet to be performed.

In addition to the effects on the activity statistics, we demonstrate that the loss of synapses results in a reduction of the network's sensitivity to small perturbations, which goes hand in hand with an increase in linear stability. In the presence of unlimited firing rate homeostasis, the perturbation sensitivity, as well as all other dynamical network characteristics are preserved, even if the extent of synapse loss is substantial. In addition to the dynamical features, the total synaptic contact area, which is decreased in the AD network due to synapse loss, is largely retained. If the homeostatic synapse growth is limited, the network dynamics as well as the total synaptic area are preserved as long as the firing rate can be maintained. Beyond this point, the network quickly approaches the state of the pathological AD network without homeostasis. The effectiveness of homeostatic compensation investigated in this study provides a possible explanation for why morphological disease-related changes in the brain (e.g. synapse loss) precede any clinically recognizable cognitive deficits by years or even decades (Morris, 2005). That homeostasis is able to recover all network characteristics is non-trivial, because in the homeostatic network with few but strong EE synapses, the statistics of the synaptic input (mean and variance) is altered with respect to the intact reference network with many weak EE synapses.

In order to investigate this observation further, we analyze the linear stability characteristics of the network and find a unique dependency of the network's spectral radius on the network's firing rate under unlimited homeostasis. Previous theoretical studies have shown that simple recurrent neuronal networks exhibit optimal computational performance for a variety of tasks if they operate in a regime where small perturbations are neither amplified nor instantly forgotten, i.e. close to the edge of chaos (Langton, 1990; Legenstein & Maass, 2007b,a; Schrauwen et al., 2009; Dambre et al., 2012; Schuecker et al., 2017). Here, we regard the network's sensitivity to a small perturbation as an indicator of its computational performance in a broad sense. Assuming that a healthy network acts close to the edge of chaos, our results suggest that the EE-synapse loss observed in AD moves the dynamics of the network away from that point towards a less sensitive regime with stable dynamics.

This key prediction of our study can be tested experimentally in animal models by analyzing time series of recorded neuronal activity. The degree of chaoticity can be revealed by the application of metrics such as the power spectrum, autocorrelation function, fractal dimension, Lyapunov's exponent (for review see Golovko et al., 2002; Beggs & Plenz, 2003; Kriener et al., 2014), and the analysis of neuronal avalanches (Friedman et al., 2012; Kriener et al., 2014; Beggs & Plenz, 2003).

Whereas our analysis accounts for why the sensitivity to perturbation recovers under unlimited homeostasis, it is notable that the coefficient of variation and the Fano factor of the spike trains are also preserved, suggesting a relationship between the transition from the stable to the chaotic regime and these two network activity characterizations. It has previously been proposed that the transition in spiking neuronal networks from the homogeneous asynchronous state (small sensitivity to perturbation and small CV) to the heterogeneous asynchronous state (high sensitivity to perturbation and high CV) of spiking networks is equivalent to the point where analogous rate networks become chaotic (Ostojic, 2014; Wieland et al., 2015). Such a relationship would also explain our observation that the maintenance of the stability of the linearized network dynamics coincidences with the maintenance of the CV.

Our results raise the question of why a shift towards more stable dynamics would be disadvantageous for the system. A network that is insensitive to perturbation in the input is prone to fading memory (changes in the external input are fast forgotten, see, e.g., Boyd & Chua 1985; Bertschinger & Natschläger 2004). Such networks are likely to be less flexible in responding to new inputs and thus harder to train than networks with chaotic dynamics. (see FORCE learning and liquid state computing; Sussillo & Abbott, 2009; Maass et al., 2002). On the other hand, insensitivity to small perturbations makes the system less susceptible to disruption by noise and is a prerequisite for the formation of stable attractors, which have been frequently used as memory storage embodiment in neuronal networks. (e.g., Li et al., 2015). However, more recent recordings in prefrontal and association cortices revealed that single cells exhibit complex and variable dynamics with respect to stimulus representation (Jun et al., 2010), which neither supports the hypothesis of stable attractors nor points to a network dynamics in the stable regime. Computational studies that have investigated the memory capacity whilst taking heterogeneous neural dynamics into account have found that memory formation succeeds well in a chaotic regime (Pereira & Brunel, 2018; Barak et al., 2013) or with an embedding of stable subspaces in chaotic dynamics (Murray et al., 2017). In addition, the construction of associative memory based on unstable periodic orbits of chaotic

attractors has been suggested as a possible way of increasing memory capacity (Wagner & Stucki, 2002). Thus, stable attractors and dynamics are not in line with experiments and might even be disadvantageous for memory formation.

On the cognitive level, these results suggest that, as homeostatic compensation mechanisms begin to fail, the shift of dynamics towards the stable regime would cause a decrease in performance within a variety of domains. For example, deficits in memory, known to primarily affect recent experiences of the AD individual, could be accounted for by the hypothesis that chaotic dynamics are needed to form new attractors (Barak et al., 2013). In addition, very stable dynamics hinder the transition from one attractor to another, which might explain the difficulties of AD patients to perform tasks switching and dual task processing (Belleville et al., 2008; Baddeley et al., 2001). Finally, the observation that AD patients often show repetitive speech and actions (Cullen et al., 2005) might be explained by difficulties in moving away from the corresponding attractor state.

So far, only a few other studies on this abstraction level exist that investigate the relationship of the physical symptoms of Alzheimer's disease to its cognitive deficits. With respect to memory, the effect of synapse loss and compensation through maintaining the TSCA has been investigated in a associative memory model (Ruppin & Reggia, 1994; Horn et al., 1993, 1996). In accordance with our results, the impairment of memory retrieval due to (excitatory) synapse loss was shown to be successfully compensated by restoring the TSCA, if the restoration occurs sufficiently quickly. The effect of the restoration on the firing rate was not explicitly shown. Although these studies demonstrated that homeostasis via upscaling synapses can retain memory performance, they lack a systematic investigation of different network parameters and do not provide an analytical explanation for the results.

Apart from AD-related computational studies, the computational consequences of intrinsic and synaptic scaling-based, homeostasis has been investigated in previous studies as response to changes in the external input (Naudé et al., 2013; Fröhlich et al., 2008). Based on a rate network, it has been demonstrated that intrinsic homeostasis, which shifts the neurons' transfer functions, moves the network dynamics towards the chaotic regime, stabilizes network activity in the present of Hebbian synaptic plasticity and improves input separability in response to an increasing external input (Naudé et al., 2013). With respect to EE-synaptic scaling, the empirical study by Fröhlich et al. (2008) showed that if the deafferentation of the external input exceeds a certain threshold, slow periodic oscillations occur, which are also observed in several CNS disorders. It is not possible to directly compare these results to our findings, since in both studies it is changes in the external input, and not the loss of recurrent connections in the network, that triggers the homeostasis response. However, both studies are in accordance with ours on the beneficial role of the homeostasis.

We complement our numerical results by an analytical approach to gain an intuitive understanding of the mechanisms underlying the recovery of the perturbation sensitivity (and hence, computational performance) by firing rate homeostasis. To study the linear stability characteristics of the network, we apply mean-field theory, similar to the approach used by Ostojic (2014). Note that we do not claim that a loss of linear stability coincides with the transition from stable to chaotic dynamics (Engelken et al., 2015; Ostojic, 2015; Kadmon & Sompolinsky, 2015; Harish & Hansel, 2015; Schuecker et al., 2017), as observed in large

autonomous random networks of analog neurons (Sompolinsky et al., 1988). Rather, we exploit that the linear stability characteristics follow a similar trend as the perturbation sensitivity. Assessing the linear stability characteristics relies on the knowledge of the effective connection strengths, i.e. the number of excess response spikes evoked by an additional input spike in the presence of synaptic background activity. This effective connectivity can be obtained experimentally (see, e.g., Boucsein et al., 2009; London et al., 2010), or, for a specific neuron and synapse model, numerically (see, e.g., Nordlie et al., 2010; Heiberg et al., 2013, 2018). For simplified models, such as the leaky integrate-and-fire neuron studied here, it can be calculated analytically under simplifying assumptions (diffusion approximation; Fourcaud & Brunel, 2002; Schuecker et al., 2015). However, we note that the preservation of linear stability by firing-rate homeostasis is due to the approximately exponential shape of the gain function. It remains to be investigated whether our results can be generalized to other types of neurons with different gain functions. Our theoretical analysis exposes the working-point dependence of the effective weights as the essential mechanism underlying the recovery of linear stability by firing rate homeostasis: on the one hand, the upscaling of EE synaptic weights required for maintaining the firing rates contributes to a destabilization of the network dynamics. On the other hand, the increase in synaptic weights leads to an increase in the variance of the synaptic-input fluctuations, which, in turn, reduces the neurons' susceptibility to modulations in the presynaptic input, and therefore stabilizes network dynamics. Note that a similar effect has been described in (Grytskyy et al., 2013).

The present study shows that certain cognitive deficits in Alzheimer's disease may be attributed to changes in the stability characteristics of neuronal network dynamics. Its central aim is to contribute a deeper insight into the relationship between disease related alterations at the structural, the dynamical and the cognitive level. The findings of this study are however also applicable in an entirely different context: in the face of limited computational resources, neuronal network models are often downscaled by reducing the number of nodes or the number of connections while increasing their strength. This downscaling has limitations if dynamical features such as the temporal structure of correlations in the neuronal activity are to be maintained (van Albada et al., 2015). The present work demonstrates that certain functional characteristics such as the sensitivity to perturbations or the classification performance can be largely preserved, if the synaptic weights are not limited by biological constraints. This insight may be particularly relevant for cognitive-computing applications based on recurrent neuronal networks implemented in neuromorphic hardware (Furber, 2016). Here, the realization of natural-density connectivity and communication constitute a major bottleneck, whereas the strength of connections is hardly limited.

The results reported in this study are based on a model of AD where synapse loss and synaptic scaling are confined to connections between excitatory neurons (EE). The motivation for restricting our investigation to the loss of EE connections is that this appears to be a prominent feature in many cortical areas (Lacor et al., 2007; Merino-Serrais et al., 2013; Dorostkar et al., 2015). Evidence that other types of synapses are also damaged in the course of the disease has been gathered from several mouse models. For example, inhibitory synapses from neurons in the entorhinal cortex to excitatory CA1 hippocampal have been found to be selectively degenerated in AD mice (Yang et al., 2016). Our mean field theoretical results suggest that a global unspecific synapse loss affecting all types of connections (EE, EI, IE, II) leads to noticeable changes in firing rates and linear stability characteristics, but only for higher levels

of synapse loss (more than 50%; see Supplements Sec. 4.5.4). In this scenario, a recovery of firing rates by a synapse unspecific scaling of synaptic weights largely preserves the linear stability characteristics, similar to our findings obtained for a EE-synapse loss and EE-synapse scaling. This suggests that the commonly reported scaling of EE-synapses may well be a mechanism the brain employs to compensate for alterations in dynamical characteristics that are induced by other types of synapse loss.

Although synapse loss correlates best with the cognitive decline observed in AD, by focusing on this aspect, the current study neglects other physical manifestations of AD such as neuron death and alterations of intrinsic neuronal properties (Hoxha et al., 2012; Haghani et al., 2012; Liu et al., 2013; Corbett et al., 2013; Eslamizade et al., 2015). These phenomena would affect both inhibition and excitation in the network, so the changes of the resulting firing rate may well be non-monotonic, unlike in our model, having unpredictable effects on the the computational properties. Alternatively, they might be entirely unaffected: in a computational study, Barrett et al. (2016) showed that under some circumstances, a network can compensate for neuron loss without the need for additional homeostasis mechanisms by adjusting neuronal transfer functions. The contribution of intrinsic neuron contributions to the claimed hyperexcitability of inhibitory neurons observed in AD has been previously investigated in a computational study by Perez et al. (2016). Whereas the interplay of such properties with synaptic loss and homeostasis are out of scope of the current work, our model could be extended to incorporate these aspects. However, there is as yet no consensus on which cell type shows hyperactivity (Zilberter et al., 2013) or hypoactivity (Yun et al., 2006); which moreover may vary over the course of the disease (Busche et al., 2012; Orbán et al., 2010).

Analogously to our focus on synaptic loss to EE connections, we also restricted our investigation of firing rate homeostasis to EE-synapse growth. This is motivated by the findings that intense synaptic upscaling is observed in AD and that an increase of excitatory-excitatory connections has been reported as a main compensation mechanism that increases the firing rate in hippocampal and cortical neurons after an artificially induced decrease in activity (e.g. by blocking sodium channels (TTX) or glutamatergic synapses or AMPAR) (Lissin et al., 1998; O'Brien et al., 1998; Turrigiano et al., 1998; Watt et al., 2000; Thiagarajan et al., 2005; Ibata et al., 2008; Kim & Tsien, 2008). Other mechanisms that increase the network's firing rate could also be considered, e.g. changes in current flow of ions (e.g. Desai et al., 1999; Gibson et al., 2006) or moving the spike-initiation zone (Grubb & Burrone, 2010). In order to understand the complexity of Alzheimer's disease it is important to study the effects of the different observed morphological alterations caused by AD, their corresponding homeostatic responses and, crucially, how they interfere with each other.

The findings of our study suggest that homeostatic synaptic scaling might be an attractive target for drug development. However, some caution is required. Firstly, as discussed above, during early AD the neuronal activity seems to be increased, followed by a decrease. Thus, enhancing EE-synaptic scaling at the very beginning of AD manifestation could even accelerate the progression of the disease. In the later stages of the disease, supporting synaptic scaling might be beneficial, stabilizing the cognitive performance. Within this context, there are a variety of molecular substrates that regulate synaptic scaling, and which show altered expression patterns in AD, that could be considered as treatment targets, for example MSK1,

PSD-95, BDNF, Arc, Calcineurin, CaMK4 and Cdk5 (for reviews see Jang & Chung 2016). A major challenge is to determine whether the altered concentrations of these substrates are a consequence of direct AD pathology, or arise as an attempt of the organism to counteract pathology, or even a mixture of both. Thus, in addition to more comprehensive modelling investigations, further research on the exact time line of morphological changes and their functional implications is needed to identify promising therapeutic targets.

## 4.4 Methods

### 4.4.1 Network model

The network consists of $N = N_\mathrm{E} + N_\mathrm{I}$ identical leaky integrate-and-fire neurons, subdivided into a population of $N_\mathrm{E} = 1000$ excitatory and a population of $N_\mathrm{I} = N_\mathrm{E}/4$ inhibitory neurons. In the intact reference network, each excitatory (inhibitory) neuron receives local excitatory inputs from $K_\mathrm{EE} = \epsilon N_\mathrm{E}$ ($K_\mathrm{IE} = \epsilon N_\mathrm{E}$) randomly selected excitatory neurons, and inhibitory inputs from $K_\mathrm{EI} = \epsilon N_\mathrm{I}$ ($K_\mathrm{II} = \epsilon N_\mathrm{I}$) randomly selected inhibitory neurons. In addition, the neurons in the local circuit are driven by external excitatory inputs modeled as an ensemble of $p$ Poissonian spike trains with constant rate $\nu_\mathrm{X}$. Each of these external spike trains is sent to a subset of $K_\mathrm{X}^\mathrm{out}$ randomly selected (excitatory and inhibitory) neurons in the network. Synaptic interactions are implemented in the form of stereotype exponential postsynaptic currents with a time constant $\tau_\mathrm{s}$. The strength $J_{ij}$ of interaction between two neurons $j$ and $i$, the synaptic weight, is parameterized by the amplitude of the postsynaptic potential of neuron $i$ evoked by an incoming spike from neuron $j$. In the reference network, all excitatory connections and all inhibitory connections, respectively, have equal synaptic weights, i.e. $J_\mathrm{EE} = J_\mathrm{IE} = J$ and $J_\mathrm{EI} = J_\mathrm{II} = -gJ$. The greater number of excitatory inputs is compensated by a larger amplitude of inhibitory synaptic weights ($g = 6$).

AD is implemented by a systematic removal of excitatory synapses to excitatory neuron (EE synapses), i.e. by a reduction in the in-degree $K_\mathrm{EE}$. All other in-degrees ($K_\mathrm{IE}$, $K_\mathrm{EI}$, $K_\mathrm{II}$) are preserved. In the presence of firing-rate homeostasis, the removal of EE connections is compensated by increasing the weights $J_\mathrm{EE}$ of the remaining EE synapses such that the time and population averaged firing rate $\nu = (NT)^{-1} \sum_{i=1}^{N} \int_0^T dt\, s_i(t)$ is preserved. Here, $s_i(t)$ denotes the spike train generated by neuron $i$ (see below), and $T = 1\,\mathrm{s}$ the simulation time. The upscaling of the EE weights $J_\mathrm{EE}$ is performed through bisectioning with an initial weight increment $\Delta J_\mathrm{EE} = J_\mathrm{EE}$. The algorithm is stopped once the population averaged firing rate $\nu$ matches the rate of the corresponding intact reference network up to a precision of $0.5\%$. In the case of limited homeostasis, $J_\mathrm{EE}$ is set to $1.2J$ if the solution of the bisectioning exceeds $120\%$ of the reference weight $J$. The weights $J_\mathrm{IE}$, $J_\mathrm{EI}$ and $J_\mathrm{II}$ of all other connections are not changed by the firing-rate homeostasis.

Unless stated otherwise, the network simulations are repeated for $M = 10$ random realizations of network connectivity, initial conditions and external inputs for each parameter configuration. A detailed description of the network model components, dynamics and

parameters is given in the Supplementary Material (Sec. 4.5.1 and Sec. 4.5.2). Simulations were performed using NEST version 2.10.0.

### 4.4.2 Synaptic contact area and characterization of network activity

**Relative total synaptic contact area** We calculate the total synaptic contact area (TSCA) of the EE synapses as the product $J_{\mathrm{EE}}K_{\mathrm{EE}}$ of the EE weight $J_{\mathrm{EE}}$ and the EE in-degree $K_{\mathrm{EE}}$. The

$$\text{relative TSCA} = \frac{K_{\mathrm{EE}}J_{\mathrm{EE}}}{{K_{\mathrm{EE}}}^{\mathrm{ref}}J_{\mathrm{EE}}^{\mathrm{ref}}} \tag{4.1}$$

is given by the ratio of the TSCA of the neurodegenerated network (reduced in-degree $K_{\mathrm{EE}}$) and the TSCA of the corresponding intact reference network (full in-degree $K_{\mathrm{EE}}$) with identical weights $J_{\mathrm{IE}}$, $J_{\mathrm{EI}}$ and $J_{\mathrm{II}}$.

**Spiking activity** We represent the spike train $s_i(t) = \sum_k \delta(t - t_{i,k})$ of neuron $i$ ($i \in [1, N]$) as the superposition of Dirac-delta functions centered about the spike times $t_{i,k}$ ($k = 1, 2, \ldots$). The spike count $n_i(t; b)$ is given by the number of spikes emitted in the time interval $[t, t + b]$. For subsequent analyses, we further compute the low-pass filtered spiking activity $x_i(t) = (s_i * h)(t)$ of neuron $i$ as the linear convolution of its spike train $s_i(t)$ with an exponential kernel $h(t) = \exp(-t/\tau_{\mathrm{f}})\,\Theta(t)$ with time constant $\tau_{\mathrm{f}}$ and Heaviside step function $\Theta(t)$.

**Average firing rate** The time and population averaged firing rate $\nu = (NT)^{-1}\sum_{i=1}^{N} n_i(0; T)$ is given by the total number $\sum_{i=1}^{N} n_i(T)$ of spikes emitted in the time interval $[0, T]$, normalized by the network size $N$ and the observation time $T = 10\,\mathrm{s}$.

**Fano factor** As a global measure of spiking synchrony, we employ the Fano factor

$$\mathrm{FF}(b) = \frac{\mathrm{Var}_t(n(t; b))}{\langle n(t; b)\rangle_t} \tag{4.2}$$

of the population spike count $n(t; b) = \sum_{i=1}^{N} n_i(t; b)$ for a binsize $b = 10\,\mathrm{ms}$. $\langle n(t; b)\rangle_t$ and $\mathrm{Var}_t(n(t; b)$ denote the mean and the variance of the population spike count $n(t; b)$ across time, respectively. Here, we exploit the fact that the variance of a sum signal $n(t)$ is dominated by pairwise correlations between the individual components $n_i(t)$, if the number $N$ of components is large (see, e.g., Harris & Thiele, 2011; Tetzlaff et al., 2012). Normalization by the mean $\langle n(t; b)\rangle_t$ ensures that $\mathrm{FF}(b)$ does not trivially depend on the firing rate or the binsize $b$. For an ensemble of $N$ independent realizations of a stationary Poisson process, $\mathrm{FF}(b) = 1$, irrespective of $b$ and the firing rate. In this work, an increase in FF indicates an increase in synchrony on a time scale $b$.

**Coefficient of variation** The degree of spiking irregularity of neuron $i$ is quantified by the coefficient of variation $\mathrm{CV}_i = \mathrm{SD}_k(\tau_{i,k}) / \langle \tau_{i,k} \rangle_k$ of the inter-spike intervals $\tau_{i,k} = t_{i,k} - t_{i,k-1}$, i.e. the ratio between the standard deviation $\mathrm{SD}_k(\tau_{i,k})$ and the mean $\langle \tau_{i,k} \rangle_k$. For a stationary Poisson point process, $\mathrm{CV}_i = 1$, irrespective of its firing rate. CV's larger (smaller) than 1 correspond to spike trains that are more (less) regular than a stationary Poisson process. We measure $\mathrm{CV}_i$ over a time interval $T = 10\,\mathrm{s}$, and report the population average $\mathrm{CV} = N^{-1} \sum_{i=1}^{N} \mathrm{CV}_i$.

**Sensitivity to perturbation** We examine the sensitivity of a network to a small perturbation in the input spikes by performing two simulations with identical initial conditions and identical realizations of external inputs. In the second run, we apply a small perturbation by delaying one spike in one external Poisson input at time $t^* = 400\,\mathrm{ms}$ by $\delta t^* = 0.5\,\mathrm{ms}$. As a measure of the network's perturbation sensitivity, we compute the Pearson correlation coefficient

$$R(t) = \frac{\langle \delta x_i(t) \delta x_i^*(t) \rangle_i}{\sqrt{\langle \delta x_i(t)^2 \rangle_i \langle \delta x_i^*(t)^2 \rangle_i}} \tag{4.3}$$

of the low-pass filtered spike responses $x_i(t)$ and $x_i^*(t)$ in the unperturbed and perturbed simulation, respectively, for each time point $t$. Here, $\delta x_i(t) = x_i(t) - \langle x_i(t) \rangle_i$ denotes the deviation of the low-pass filtered spike response $x_i(t)$ of neuron $i$ from the population average $\langle x_i(t) \rangle_i$. $\langle \ldots \rangle_i = N^{-1} \sum_{i=1}^{N} \ldots$ represents the population average. We define the time-dependent and the long-term perturbation sensitivity as $S(t) = 1 - |R(t)|$ (Fig. 4.5, bottom panels) and $S = S(t_{\mathrm{obs}} = 10\,\mathrm{s})$ (Fig. 4.6), respectively. An observation of $S(t_{\mathrm{obs}}) = 0$ indicates that the effect of the small perturbation has vanished, i.e. that the network has stable dynamics and is insensitive to the perturbation. An observation of $S(t_{\mathrm{obs}}) = 1$, in contrast, corresponds to diverging spike patterns in response to the perturbation and thus chaotic dynamics. In dynamical-systems theory and related applications, state differences are typically expressed in terms of the Euclidean distance $D = \sqrt{\sum_{i=1}^{N} [x_i(t) - x_i^*(t)]^2}$. Here, we employ the (normalized) correlation coefficient $R$ instead to avoid (trivial) firing rate dependencies. Note that $D$ and $R$ are redundant in the sense that both can be expressed in terms of the moments $\langle x_i(t) x_i^*(t) \rangle_i$, $\langle x_i(t)^2 \rangle_i$, $\langle x_i^*(t)^2 \rangle_i$, $\langle x_i \rangle_i$ and $\langle x_i^* \rangle_i$.

### 4.4.3 Linearized network dynamics and stability analysis

In the following, we describe the analytical approach to investigate the effect of synapse loss and firing rate homeostasis on the network's linear-stability characteristics. To this end, we employ results obtained from the diffusion approximation of the leaky-integrate-and-fire (LIF) neuron with exponential postsynaptic currents under the assumption that the synaptic time constant $\tau_{\mathrm{s}}$ is small compared to the membrane time constant $\tau_{\mathrm{m}}$, and that the network activity is sufficiently asynchronous and irregular (mean-field theory; Fourcaud & Brunel, 2002; Helias et al., 2013; Schuecker et al., 2015). All parameters that are not explicitly mentioned here can be found in tab. 4.2.

**Stationary firing rates and fixed points** For each parameter set (synaptic weight $J$, extent of synapse loss, different types of firing rate homeostasis), we first identify the self-consistent stationary states by solving

$$\begin{aligned}\nu_\mathrm{E} &= G\left(\mu_\mathrm{E}(\nu_\mathrm{E},\nu_\mathrm{I}), \sigma_\mathrm{E}(\nu_\mathrm{E},\nu_\mathrm{I})\right)\\ \nu_\mathrm{I} &= G\left(\mu_\mathrm{I}(\nu_\mathrm{E},\nu_\mathrm{I}), \sigma_\mathrm{I}(\nu_\mathrm{E},\nu_\mathrm{I})\right)\end{aligned} \tag{4.4}$$

for the population averaged firing rates $\nu_\mathrm{E}$ and $\nu_\mathrm{I}$ of the excitatory and inhibitory subpopulations. Here,

$$G(\mu,\sigma) = \left(\tau_\mathrm{ref} + \tau_\mathrm{m}\sqrt{\pi}\int_{y_\mathrm{r}}^{y_\theta} du\, f(u)\right)^{-1} \tag{4.5}$$

represents the stationary firing rate of the LIF neuron in response to a synaptic input current with mean $\mu$ and variance $\sigma^2$ in diffusion approximation, with $f(u) = e^{u^2}\left[1+\mathrm{erf}(u)\right]$, $y_\mathrm{r} = (V_\mathrm{r}-\mu)/\sigma + \frac{q}{2}\sqrt{\tau_\mathrm{s}/\tau_\mathrm{m}}$, $y_\theta = (\theta-\mu)/\sigma + \frac{q}{2}\sqrt{\tau_\mathrm{s}/\tau_\mathrm{m}}$ and $q = \sqrt{2}|\zeta(1/2)|$ (with Riemann zeta function $\zeta$; Fourcaud & Brunel, 2002; Helias et al., 2013; Schuecker et al., 2015). For stationary firing rates $\nu_\mathrm{E}$ and $\nu_\mathrm{I}$ of the local presynaptic neurons, the mean and the variances of the total synaptic input currents to excitatory and inhibitory neurons are given by

$$\begin{aligned}\mu_\mathrm{E} &= \left(K_\mathrm{EE}\hat{J}_\mathrm{EE}\nu_\mathrm{E} + K_\mathrm{EI}\hat{J}_\mathrm{EI}\nu_\mathrm{I} + K_\mathrm{X}\hat{J}_\mathrm{X}\nu_\mathrm{X}\right)\tau_\mathrm{m},\\ \mu_\mathrm{I} &= \left(K_\mathrm{IE}\hat{J}_\mathrm{IE}\nu_\mathrm{E} + K_\mathrm{II}\hat{J}_\mathrm{II}\nu_\mathrm{I} + K_\mathrm{X}\hat{J}_\mathrm{X}\nu_\mathrm{X}\right)\tau_\mathrm{m},\\ \sigma_\mathrm{E}^2 &= \left(K_\mathrm{EE}\hat{J}_\mathrm{EE}^2\nu_\mathrm{E} + K_\mathrm{EI}\hat{J}_\mathrm{EI}^2\nu_\mathrm{I} + K_\mathrm{X}\hat{J}_\mathrm{X}^2\nu_\mathrm{X}\right)\tau_\mathrm{m},\\ \sigma_\mathrm{I}^2 &= \left(K_\mathrm{IE}\hat{J}_\mathrm{IE}^2\nu_\mathrm{E} + K_\mathrm{II}\hat{J}_\mathrm{II}^2\nu_\mathrm{I} + K_\mathrm{X}\hat{J}_\mathrm{X}^2\nu_\mathrm{X}\right)\tau_\mathrm{m},\end{aligned} \tag{4.6}$$

respectively. The coefficients $K_{pq}$ ($p,q \in \{\mathrm{E},\mathrm{I}\}$) denote the number of inputs (in-degree) to neurons in population $p$ from population $q$, $\hat{J}_{pq} = \tau_\mathrm{s}C_\mathrm{m}^{-1}\hat{I}_{pq}$ the corresponding rescaled PSC amplitude, $K_\mathrm{X}$ the number of external inputs for each neuron in the network, and $\nu_\mathrm{X}$ the firing rate of the Poissonian external sources. Note that in our network simulations, each external source is connected to a randomly selected subset of $K_\mathrm{X}^\mathrm{out}$ neurons. As a result, the number $K_\mathrm{X}$ of external inputs each neuron in the network receives is a binomially distributed random number. For the analytical treatment, we neglect this variability and replace $K_\mathrm{X}$ by the average $K_\mathrm{X} = pK_\mathrm{X}^\mathrm{out}/N$. Equations eq. 4.4 and eq. 4.6 are simultaneously solved numerically using the `optimize.root()` function (`method='hybr'`) of the `scipy` package (`http://www.scipy.org`). To ensure that all solutions are found, the fixed-point search is repeated for 30 pairs of initial rates randomly drawn from a uniform distribution between 0 and 50 spikes/s. If multiple coexisting fixed points are found, the one with the highest firing rates is chosen for the subsequent analysis.

**Synapse loss and firing rate homeostasis** In this work, Alzheimer's disease is modeled by removing a fraction of EE synapses, i.e. by reducing the in-degree $K_\mathrm{EE}$. The self-consistent firing rates $\nu_\mathrm{E}$ and $\nu_\mathrm{I}$ after synapse removal are hence reduced (Fig. 4.8 D). In the presence of unlimited firing rate homeostasis, we adjust the weight $\hat{J}_\mathrm{EE}$ (of the remaining synapses) (Fig. 4.8 B) until the excitatory self-consistent firing rate $\nu_\mathrm{E}^\mathrm{ref}$ of the intact reference network (before synapse removal) is recovered (Fig. 4.8 E). To this end, we numerically find the roots

of $\nu_E - \nu_E^{ref}$ by employing again `scipy`'s `optimize.root()` function. We repeat the root finding for 30 initial weights randomly drawn from a uniform distribution between $\hat{J}_{EE}^{ref}$ and $10\hat{J}_{EE}^{ref}$, where $\hat{J}_{EE}^{ref}$ denotes the original weight before synapse removal, and keep the solution where $|\nu_E - \nu_E^{ref}|$ is minimal. For limited homeostasis, the new EE weight is chosen as the minimum of the solution $\hat{J}_{EE}$ and $1.2\hat{J}_{EE}^{ref}$ (Fig. 4.8 C,F).

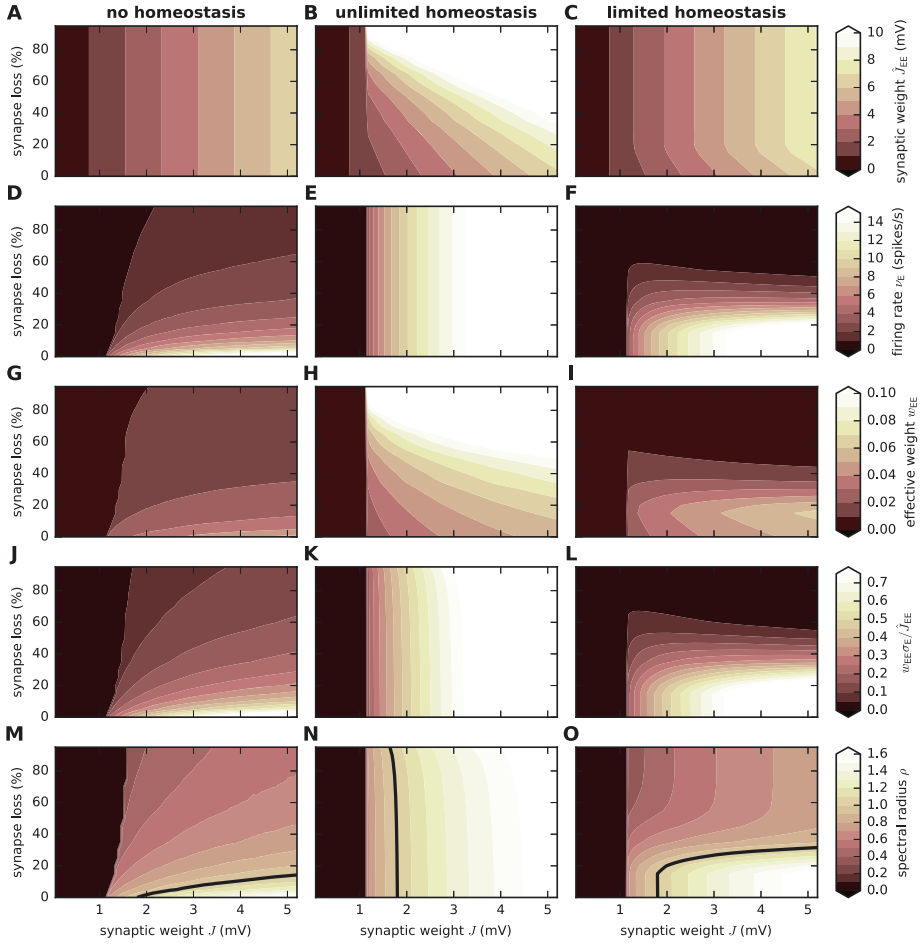


**Figure 4.8: Mean-field theory.** Dependence of the synaptic weight $\hat{J}_{EE}$ (**A**–**C**), the average firing rate $\nu_E$ of the excitatory population (**D**–**F**), the effective weight $w_{EE}$ of EE connections (**G**–**I**), the ratio $w_{EE}\sigma_E/\hat{J}_{EE}$ (**J**–**L**), and the spectral radius $\rho$ (**M**–**O**) on the synaptic weight $J$ and the degree of synapse loss in the absence of homeostatic compensation (left column), as well as with unlimited (middle column) and limited firing rate homeostasis (right column). Superimposed black curves in (M–O) mark instability lines $\rho = 1$. Same parameters as in network simulations (see Sec. 4.5.1 and Sec. 4.5.2).

**Linearized network dynamics and effective connectivity** As shown in (Tetzlaff et al., 2012; Helias et al., 2013), networks of spiking neurons can be formally linearized about a stationary state $\boldsymbol{\nu}^* = (\nu_1^*, \ldots, \nu_N^*)$ (linear-response theory) and thereby be mapped to an $N$-dimensional system

$$\delta\nu_i(t) = \sum_{j=1}^{N} (h_{ij} * \delta\nu_j)(t) \qquad (i \in [1, N]) \tag{4.7}$$

of linear equations describing the dynamics of small firing rate fluctuations $\delta\nu_i(t) = \nu_i(t) - \nu_i^*$ around this stationary state. The stationary states are determined as the self-consistent solutions of

$$\boldsymbol{\nu}^* = \boldsymbol{\phi}(\boldsymbol{\nu}^*), \tag{4.8}$$

where $\boldsymbol{\phi}(\boldsymbol{\nu}_{\text{in}})$ represents the activation function mapping the vector of stationary input rates $\boldsymbol{\nu}_{\text{in}}$ to the vector of output rates. The coupling kernel $h_{ij}(t)$ represents the firing rate impulse response, i.e. the modulation in the output rate $\nu_i(t)$ in response to a delta-shaped fluctuation in the rate $\nu_j(t)$ of presynaptic neuron $j$. We refer to the area

$$w_{ij} = \int_{-\infty}^{\infty} dt\, h_{ij}(t) \tag{4.9}$$

under the coupling kernel as the effective connection weight. It measures the average number of extra spikes emitted by target neuron $i$ in response to a spike fired by the presynaptic neuron $j$, in the context of the background activity determined by the stationary state $\boldsymbol{\nu}^*$. Exploiting the fact that the integral of the impulse response of a linear(ized) system is identical to the long-term limit of its step response, the effective weight

$$w_{ij} = \left.\frac{\partial\phi_i(\boldsymbol{\nu})}{\partial\nu_j}\right|_{\boldsymbol{\nu}^*} \tag{4.10}$$

is given by the derivative of the activation function $\phi_i$ of neuron $i$ with respect to the stationary firing rate $\nu_j$ of neuron $j$, evaluated at the stationary state $\boldsymbol{\nu}^*$. With $\phi_i(\boldsymbol{\nu}) = G(\mu_i(\boldsymbol{\nu}), \sigma_i(\boldsymbol{\nu}))$ from eq. 4.5, $\mu_i(\boldsymbol{\nu}) = \left(\sum_{j=1}^{N} \hat{J}_{ij}\nu_j + K_{\text{X}}\hat{J}_{\text{X}}\nu_{\text{X}}\right)\tau_{\text{m}}$, and $\sigma_i^2(\boldsymbol{\nu}) = \left(\sum_{j=1}^{N} \hat{J}_{ij}^2\nu_j + K_{\text{X}}\hat{J}_{\text{X}}^2\nu_{\text{X}}\right)\tau_{\text{m}}$, we obtain

$$w_{ij} = \frac{\partial G}{\partial\mu_i}\frac{\partial\mu_i}{\partial\nu_j}\bigg|_{\boldsymbol{\nu}^*} + \frac{\partial G}{\partial\sigma_i}\frac{\partial\sigma_i}{\partial\nu_j}\bigg|_{\boldsymbol{\nu}^*} = \frac{\hat{J}_{ij}}{\sigma_i^*}\sqrt{\pi}(\tau_{\text{m}}\nu_i^*)^2\left(f(y_{\theta i}^*) - f(y_{\text{r}i}^*)\right) \tag{4.11}$$

as the effective weight of the LIF neuron in the stationary self-consistent state given by $\boldsymbol{\nu}^*$ (Tetzlaff et al., 2012; Helias et al., 2013). Note that for the result on the right-hand side of eq. 4.11, we account only for the derivative $\frac{\partial G}{\partial\mu_i}$ of $G$ with respect to the mean input $\mu_i$ (DC susceptibility), but neglect the contribution $\frac{\partial G}{\partial\sigma_i}$ resulting from a modulation in the input variance $\sigma_i^2$. Removal of EE synapses and the resulting decrease in stationary firing rates (Fig. 4.8 D) leads to a reduction in the effective weight $w_{\text{EE}}$ of EE connections (Fig. 4.8 G). In the presence of (unlimited) firing rate homeostasis, upscaling of EE synapses (Fig. 4.8 B) and the resulting preservation of firing rates (Fig. 4.8 E) results in an increase in $w_{\text{EE}}$ (Fig. 4.8 H).

**Stability analysis** For the LIF neuron with weak exponential synapses (Helias et al., 2013) as well as for a variety of other neuron and synapse models (Nordlie et al., 2010; Heiberg et al.,

2013, 2018), the effective coupling kernel $h_{ij}(t)$ introduced in eq. 4.7 can be well approximated by an exponential function $h_{ij}(t) = w_{ij}\tau^{-1}\exp(-t/\tau)\Theta(t)$ with an effective time constant $\tau$ and Heaviside function $\Theta(t)$. With this approximation, eq. 4.7 can be written in form of an $N$-dimensional system of differential equations

$$\tau\frac{d\delta\boldsymbol{\nu}}{dt} = -\delta\boldsymbol{\nu} + \boldsymbol{W}\delta\boldsymbol{\nu}(t). \tag{4.12}$$

Here, $\boldsymbol{W} = \{w_{ij}\}$ denotes the $N \times N$ effective connectivity matrix and $\delta\boldsymbol{\nu}(t) = (\delta\nu_1(t), \ldots, \delta\nu_N(t))$ the vector of firing rate fluctuations. The system eq. 4.12 has bounded solutions only if the real parts of all Eigenvalues $\lambda_k$ of the effective connectivity matrix $\boldsymbol{W}$ are smaller than unity, i.e. if $\mathrm{Re}(\lambda_k) < 1\ (\forall k)$. If $\rho = \max_k(\mathrm{Re}(\lambda_k)) > 1$, the linearized system is unstable and fluctuations diverge. In the original nonlinear LIF network, an unbounded growth of fluctuations is prevented by the nonlinearities of the single-neuron dynamics. For large random networks where the statistics of the coupling strengths does not depend on the target nodes, the bulk of Eigenvalues $\{\lambda_k | k \in [1, N]\}$ of $\boldsymbol{W}$ is located in the complex plane within a circle centered at the coordinate origin and a radius $\rho$ which is determined by the variances of the effective connectivity (Rajan & Abbott, 2006). A single outlier is given by the Eigenvalue $\lambda_{k^*}$ associated with the Eigenvector $\boldsymbol{u}_{k^*} = (1, 1, \ldots, 1, 1)^{\mathrm{T}}$, which is given by the mean effective weight. In inhibition dominated networks, the mean synaptic weight and, hence, $\lambda_{k^*}$ are negative. The stability behaviour is therefore solely determined by the spectral radius $\rho$. For a random network composed of $N_{\mathrm{E}}$ excitatory ($j \in \mathcal{E}$; $N_{\mathrm{E}} = |\mathcal{E}|$) and $N_{\mathrm{I}}$ inhibitory neurons ($j \in \mathcal{I}$; $N_{\mathrm{I}} = |\mathcal{I}|$) with homogeneous in-degrees $K_{pq}$ ($p, q \in \{\mathrm{E}, \mathrm{I}\}$) and weights

$$w_{ij} = \begin{cases} w_{\mathrm{EE}} & \forall i \in \mathcal{E},\ j \in \mathcal{E},\ \text{connection } j \to i \text{ exists with probability } \frac{K_{\mathrm{EE}}N_{\mathrm{E}}}{NN_{\mathrm{E}}} \\ w_{\mathrm{EI}} & \forall i \in \mathcal{E},\ j \in \mathcal{I},\ \text{connection } j \to i \text{ exists with probability } \frac{K_{\mathrm{EI}}N_{\mathrm{E}}}{NN_{\mathrm{I}}} \\ w_{\mathrm{IE}} & \forall i \in \mathcal{I},\ j \in \mathcal{E},\ \text{connection } j \to i \text{ exists with probability } \frac{K_{\mathrm{IE}}N_{\mathrm{I}}}{NN_{\mathrm{E}}} \\ w_{\mathrm{II}} & \forall i \in \mathcal{I},\ j \in \mathcal{I},\ \text{connection } j \to i \text{ exists with probability } \frac{K_{\mathrm{II}}N_{\mathrm{I}}}{NN_{\mathrm{I}}} \\ 0 & \forall i, j,\ \text{connection } j \to i \text{ does not exist} \end{cases}, \tag{4.13}$$

the squared spectral radius is given by

$$\rho^2 = N_{\mathrm{E}}v_{\mathrm{E}} + N_{\mathrm{I}}v_{\mathrm{I}} = N^{-1}\left(K_{\mathrm{EE}}N_{\mathrm{E}}w_{\mathrm{EE}}^2 + K_{\mathrm{IE}}N_{\mathrm{I}}w_{\mathrm{IE}}^2 + K_{\mathrm{EI}}N_{\mathrm{E}}w_{\mathrm{EI}}^2 + K_{\mathrm{II}}N_{\mathrm{I}}w_{\mathrm{II}}^2\right). \tag{4.14}$$

Here, $v_{\mathrm{E}} = w_{\mathrm{EE}}^2K_{\mathrm{EE}}N_{\mathrm{E}}/(NN_{\mathrm{E}}) + w_{\mathrm{IE}}^2K_{\mathrm{IE}}N_{\mathrm{I}}/(NN_{\mathrm{I}})$ and $v_{\mathrm{I}} = w_{\mathrm{EI}}^2K_{\mathrm{EI}}N_{\mathrm{E}}/(NN_{\mathrm{I}}) + w_{\mathrm{II}}^2K_{\mathrm{II}}N_{\mathrm{I}}/(NN_{\mathrm{I}})$ denote the variances of the effective connectivity $w_{ij}$ across the ensemble of target cells ($i \in [1, N]$) for excitatory ($j \in \mathcal{E}$) and inhibitory sources ($j \in \mathcal{I}$), respectively. Without homeostatic compensation, EE-synapse loss leads to a stabilization of the linearized network dynamics, i.e. a decrease in $\rho$ (Fig. 4.8 M). In the presence of unlimited firing rate homeostasis, the spectral radius $\rho$ is preserved (Fig. 4.8 N), even if a substantial fraction of EE synapses is removed (Fig. 4.8 N). If the homeostatic resources are limited, $\rho$ is maintained until the upscaled synaptic weights reach their maximum value (Fig. 4.8 O).

**Preservation of linear stability by firing rate homeostasis** At first glance, it is unclear why firing rate homeostasis preserves the linear stability characteristics as measured by the spectral radius $\rho$. While the stationary firing rates $\nu_i^*$ are, by definition, kept constant

during synapse loss and homeostasis, the input statistics $\mu_i^*$, $\sigma_i^*$ (Fig. 4.10 D,E), $y_{r_i}^*$ and $y_{\theta i}^*$ (Fig. 4.9 A,B) as well as the effective weights $w_{ij}$ (Fig. 4.8 K) are not. To shed light on the mechanisms leading to the preservation of $\rho$, we first note that the factor

$$\sqrt{\pi}(\tau_\mathrm{m}\nu_i^*)^2\left(f(y_{\theta i}^*) - f(y_{r_i}^*)\right) =: \eta(\nu_i^*) \tag{4.15}$$

on the right-hand side of eq. 4.11 is in good approximation uniquely determined by the stationary firing rate $\nu_i^*$ (Fig. 4.8 J–L). This can be understood by noting that, according to eq. 4.5, the firing rates are determined by $\int_{y_{r_i}^*}^{y_{\theta i}^*} dy\, f(y)$, and that $f(y) = e^{y^2}[1 + \mathrm{erf}(y)]$ can be approximated by an exponential function $f(y) \approx Ae^{By}$ for the range of arguments spanned by $y_{r_i}^*$ and $y_{\theta i}^*$ (Fig. 4.9). With this approximation, $\int_{y_{r_i}^*}^{y_{\theta i}^*} dy\, f(y) = B^{-1}\left[f(y_{\theta i}^*) - f(y_{r_i}^*)\right]$. For constant firing rate, $f(y_{\theta i}^*) - f(y_{r_i}^*)$ is therefore constant, too, and the effective weight is essentially determined by the ratio $\hat{J}_{ij}/\sigma_i^*$. With $w_{pq} = \eta(\nu_p^*)\hat{J}_{pq}/\sigma_p^*$ ($p, q \in \{\mathrm{E}, \mathrm{I}\}$), eq. 4.14 reads

$$\begin{aligned}\rho^2 &= N^{-1}\left(K_{\mathrm{EE}}N_\mathrm{E}\frac{\hat{J}_{\mathrm{EE}}^2}{\sigma_\mathrm{E}^{*2}}\eta^2(\nu_\mathrm{E}^*) + K_{\mathrm{IE}}N_\mathrm{I}\frac{\hat{J}_{\mathrm{IE}}^2}{\sigma_\mathrm{I}^{*2}}\eta^2(\nu_\mathrm{I}^*) + K_{\mathrm{EI}}N_\mathrm{E}\frac{\hat{J}_{\mathrm{EI}}^2}{\sigma_\mathrm{E}^{*2}}\eta^2(\nu_\mathrm{E}^*) + K_{\mathrm{II}}N_\mathrm{I}\frac{\hat{J}_{\mathrm{II}}^2}{\sigma_\mathrm{I}^{*2}}\eta^2(\nu_\mathrm{I}^*)\right) \\ &= N^{-1}\left(\eta^2(\nu_\mathrm{E}^*)N_\mathrm{E}\frac{K_{\mathrm{EE}}\hat{J}_{\mathrm{EE}}^2 + K_{\mathrm{EI}}\hat{J}_{\mathrm{EI}}^2}{\sigma_\mathrm{E}^{*2}} + \eta^2(\nu_\mathrm{I}^*)N_\mathrm{I}\frac{K_{\mathrm{IE}}\hat{J}_{\mathrm{IE}}^2 + K_{\mathrm{II}}\hat{J}_{\mathrm{II}}^2}{\sigma_\mathrm{I}^{*2}}\right).\end{aligned} \tag{4.16}$$

According to our network simulations as well as the mean-field theory described above, stationary firing rates of the excitatory and inhibitory subpopulation are identical in the presence of firing rate homeostasis, i.e. $\nu^* := \nu_\mathrm{E}^* = \nu_\mathrm{I}^*$. With Eq.4.6 and assuming that the contribution $K_\mathrm{X}\hat{J}_\mathrm{X}^2\nu_\mathrm{X}$ of the external drive to the total input variances $\sigma_{\mathrm{E/I}}^{*2}$ can be neglected (which is the case for the range of parameters considered in this study), we find that the spectral radius

$$\rho^2 = \frac{\eta^2(\nu^*)}{\nu^*\tau_\mathrm{m}} \tag{4.17}$$

is in good approximation uniquely determined by the stationary firing rate $\nu^*$ (Fig. 4.10 and Fig. 4.7B). A constant firing rate (as achieved by firing rate homeostasis) is therefore accompanied by a constant spectral radius.

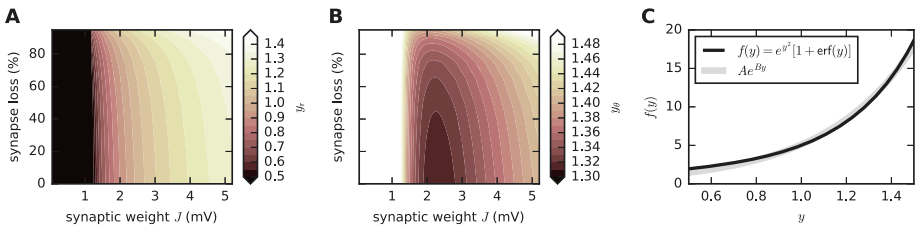


**Figure 4.9: Approximation of $f(y) = e^{y^2}[1 + \mathbf{erf}(y)]$ by an exponential function. A,B)** Dependence of $y_{r\mathsf{E}}$ (A) and $y_{\theta\mathsf{E}}$ (B) on the synaptic reference weight $J$ and the degree of synapse loss in the presence of unlimited firing rate homeostasis (mean-field theory). **C)** Graph of $f(y) = e^{y^2}[1 + \mathrm{erf}(y)]$ (black) and exponential function $Ae^{BX}$ (gray; $A = 0.4$, $B = 2.5$) fitted to $f(y)$ in interval $y \in [0.5, 1.5]$. Same parameters as in network simulations (see Sec. 4.5.1 and Sec. 4.5.2).

## 4.5 Supplementary Materials

### 4.5.1 Network model

| Summary | |
|---|---|
| **Populations** | excitatory population $\mathcal{E}$, inhibitory population $\mathcal{I}$ |
| **Connectivity** | random convergent connections (fixed in-degrees) |
| **Neuron model** | leaky integrate-and-fire (LIF) |
| **Synapse model** | exponentially decaying postsynaptic currents, static synaptic weights, fixed delays |
| **Input** | Poissonian spike trains |

| Populations | | |
|---|---|---|
| **Name** | **Elements** | **Size** |
| $\mathcal{E}$ | LIF | $N_{\mathsf{E}} = K/\epsilon$ |
| $\mathcal{I}$ | LIF | $N_{\mathsf{I}} = \gamma N_{\mathsf{E}} = \gamma K/\epsilon$ |

| Connectivity | | |
|---|---|---|
| **Source** | **Target** | **Pattern** |
| $\mathcal{E}$ | $\mathcal{E}$ | random convergent, in-degree $K_{\mathsf{EE}}$, delay $d$, weight $J_{\mathsf{EE}}$ |
| $\mathcal{E}$ | $\mathcal{I}$ | random convergent, in-degree $K_{\mathsf{IE}} = K$, delay $d$, weight $J_{\mathsf{IE}} = J$ |
| $\mathcal{I}$ | $\mathcal{E}$ | random convergent, in-degree $K_{\mathsf{EI}} = \gamma K$, delay $d$, weight $J_{\mathsf{EI}} = -gJ$ |
| $\mathcal{I}$ | $\mathcal{I}$ | random convergent, in-degree $K_{\mathsf{II}} = K$, delay $d$, weight $J_{\mathsf{II}} = -gJ$ |
| all | all | no self-connections ("autapses"), no multiple connections ("multapses") |

| Neuron | |
|---|---|
| **Type** | leaky integrate-and-fire (LIF) model |
| **Description** | dynamics of membrane potential $V_i(t)$ $(i \in \{1, \ldots, N\})$<br>•spike emission at $t_k^i$ if $V_i\left(t_k^i\right) \geq \theta$<br>•subthreshold dynamics: $\tau_{\mathsf{m}}\dot{V}_i = -V_i + R_{\mathsf{m}} I_i(t) \quad \forall k,\ \forall t \notin \left[t_k^i,\ t_k^i + \tau_{\mathsf{ref}}\right)$<br>•reset and refractoriness: $V_i(t) = V_{\mathsf{r}} \quad \forall k,\ \forall t \in \left(t_k^i,\ t_k^i + \tau_{\mathsf{ref}}\right]$<br>initial membrane-potential distribution at $t = 0$: random uniform between $0$ and $\theta$<br>exact integration with continuous spike times in discrete-time simulation (Rotter & Diesmann, 1999; Morrison et al., 2007; Hanuschkin et al., 2010)<br>temporal resolution $\Delta t$ |

| Synapse | |
|---|---|
| **Type** | current based synapses with exponential post-synaptic currents (PSCs) |
| **Description** | $I_i(t) = \sum_{j=1}^{N} \hat{I}_{ij}(\mathrm{PSC} * s_j)(t)$<br>with $\mathrm{PSC}(t) = e^{-t/\tau_\mathsf{s}}\Theta(t)$ and Heaviside function $\Theta(t) = \begin{cases} 1 & t \geqslant 0 \\ 0 & \text{else} \end{cases}$<br>$\curvearrowright$ post-synaptic potential $\mathrm{PSP}_{ij}(t) = \hat{I}_{ij}\frac{R_\mathsf{m}\tau_\mathsf{s}}{\tau_\mathsf{s}-\tau_\mathsf{m}}\left(e^{-t/\tau_\mathsf{s}} - e^{-t/\tau_\mathsf{m}}\right)\Theta(t)$<br>synaptic weight $J_{ij} = \hat{I}_{ij}\frac{R_\mathsf{m}\tau_\mathsf{s}}{\tau_\mathsf{s}-\tau_\mathsf{m}}\left(\left[\frac{\tau_\mathsf{m}}{\tau_\mathsf{s}}\right]^{\frac{-\tau_\mathsf{m}}{\tau_\mathsf{m}-\tau_\mathsf{s}}} - \left[\frac{\tau_\mathsf{m}}{\tau_\mathsf{s}}\right]^{\frac{-\tau_\mathsf{s}}{\tau_\mathsf{m}-\tau_\mathsf{s}}}\right) = \max_t(\mathrm{PSP}_{ij}(t))$ |
| **Input** | |
| **Type** | spike trains modeled as independent realizations of a Poisson point process |
| **Description** | $p$ independent Poisson spike trains of rate $\nu_\mathsf{X}$, each connected to $K_\mathsf{X}^\mathsf{out}$ randomly chosen (excitatory and inhibitory) network neurons |
| **Realizations** | |
| **Description** | repetition of network simulations for $M$ random realizations of network connectivity, initial conditions, and external inputs |

**Table 4.1:** Description of the network model according to (Nordlie et al., 2009).

### 4.5.2 Lists of parameters

| Connectivity | | |
|---|---|---|
| **Name** | **Value** | **Description** |
| $K$ | $100$ | excitatory in-degree (number of excitatory inputs) of reference network |
| $K_\mathsf{EE}$ | $5, 10, 20, 30, \ldots, 100$ | excitatory-excitatory in-degree |
| $\epsilon$ | $0.1$ | network density |
| $\gamma$ | $1/4$ | relative size of inhibitory subpopulation |
| **Neuron** | | |
| **Name** | **Value** | **Description** |
| $\tau_\mathsf{m}$ | $20\,\mathrm{ms}$ | membrane time constant |
| $\tau_\mathsf{ref}$ | $2\,\mathrm{ms}$ | absolute refractory period |
| $C_\mathsf{m}$ | $250\,\mathrm{pF}$ | membrane capacity |
| $V_\mathsf{r}$ | $0.0\,\mathrm{mV}$ | reset potential |
| $\tau_\mathsf{s}$ | $2\,\mathrm{ms}$ | time constant of post-synaptic current |
| $\theta$ | $15\,\mathrm{mV}$ | spike threshold |

| Synapse | | |
|---|---|---|
| **Name** | **Value** | **Description** |
| $J$ | $0.05, 0.1, \ldots, 4.95$ mV | EPSP amplitude |
| $g$ | $6$ | relative IPSP amplitude |
| $d$ | $1$ ms | spike transmission delay |
| **Input** | | |
| **Name** | **Value** | **Description** |
| $\nu_{\mathsf{X}}$ | $750$ spikes/s | rate of external Poisson inputs |
| $J_{\mathsf{X}}$ | $0.2$ mV | PSP amplitude evoked by external inputs |
| $p$ | 5 | number of input sources (spike trains) |
| $K_{\mathsf{X}}^{\mathsf{out}}$ | 300 | number of neurons each input source is connected to (out-degree) |
| **Simulation** | | |
| **Name** | **Value** | **Description** |
| $T$ | $0.4$, $10$ or $2000$ s | total simulation time |
| $\Delta t$ | $0.1$ ms | time resolution |
| $M$ | $10$ | number of random network realizations per parameter configuration |

**Table 4.2:** Network and simulation parameters

| Spike-train statistics | | |
|---|---|---|
| **Name** | **Value** | **Description** |
| $b$ | $10$ ms | binsize for evaluation of Fano factor |
| **Perturbation sensitivity** | | |
| **Name** | **Value** | **Description** |
| $t^*$ | $400$ ms | perturbation time |
| $\delta t^*$ | $0.5$ ms | perturbation magnitude (time shift of one input spike) |
| $t_{\mathsf{obs}}$ | $10$ s | observation time |
| $\Delta t_{\mathsf{f}}$ | $1$ ms | time resolution of low-pass filtered spiking activity |
| $\tau_{\mathsf{f}}$ | $20$ ms | time constant of low-pass filter $h(t)$ |

**Table 4.3:** Parameters for evaluation of spike-train statistics and perturbation sensitivity

### 4.5.3 Canceling of the synaptic-weight variance by the input variance

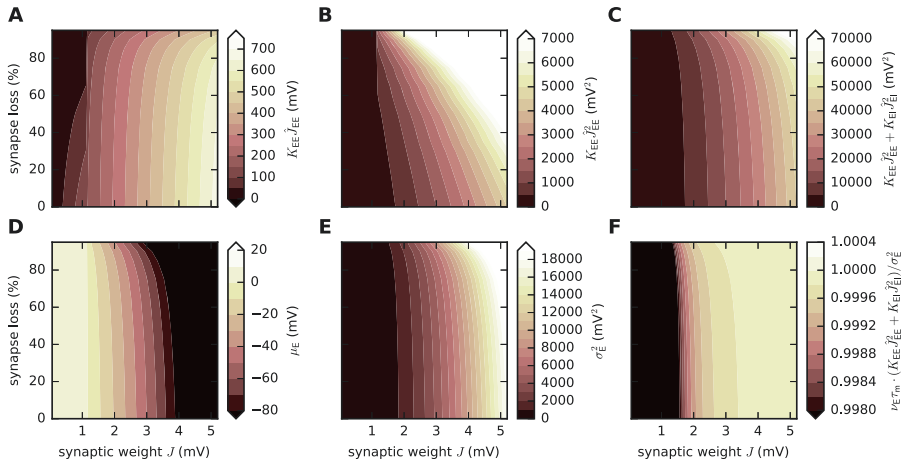


**Figure 4.10: Canceling of the synaptic-weight variance by the input variance.** Dependence of $K_{EE}\hat{J}_{EE}$ (**A**), $K_{EE}\hat{J}^2_{EE}$ (**B**), $K_{EE}\hat{J}^2_{EE}+K_{EI}\hat{J}^2_{EI}$ (**C**), input mean $\mu_E$ (**D**), input variance $\sigma^2_E$ (**E**), and the ratio $\nu_E\tau_m\,(K_{EE}\hat{J}^2_{EE}+K_{EI}\hat{J}^2_{EI})/\sigma^{*2}_E$ (**F**) on the synaptic reference weight $J$ and the degree of synapse loss in the presence of unlimited firing rate homeostasis (mean-field theory). Note that $\nu_E\tau_m\,(K_{EE}\hat{J}^2_{EE}+K_{EI}\hat{J}^2_{EI})/\sigma^{*2}_E$ (F) is very close to unity in all regions where $\nu_E > 0$ (cf.Fig. 4.8 E). Hence, the ratio between the synaptic-weight variance $K_{EE}\hat{J}^2_{EE}+K_{EI}\hat{J}^2_{EI}$ and the synaptic-input variance $\sigma^{*2}_E$ is uniquely determined by the firing rate. Same parameters as in network simulations (see Sec. 4.5.1 and Sec. 4.5.2).

### 4.5.4 Unspecific synapse loss and homeostasis

In this section, we expand our analysis of the linearized network dynamics towards a network in which all types of synapses (EE,EI,IE,II) are removed. Accordingly, the homeostatic upscaling affects all types of synapses (EE,EI,IE,II) such that the target firing rate is reached by applying the same factor $c$ to all synaptic weights and the initial proportion of the different synapse types is kept constant ($c \cdot J_{EE} = c \cdot J_{IE} = c \cdot J$ and $c \cdot J_{EI} = c \cdot J_{II} = -cg \cdot J$ with $c \geq 1$).

We observe that synapse-unspecific network dilution leads to a drop in firing rate (Fig. 4.11 D), but this drop is not as pronounced as if only EE synapses are removed (Fig. 4.8 D). For small and moderate degrees of synapse loss, the firing rate changes only little. Upscaling $J$ compensates for this and fully restores the firing rates (Fig. 4.11 E), even for high levels of synapse loss. In the absence of homeostasis, synapse unspecific network dilution reduces the spectral radius (Fig. 4.11 M), but this effect is weaker as if only EE synapses were removed (Fig. 4.8 M). For small and moderate degrees of synapse loss, the spectral radius is hardly affected. Upscaling $J$ fully recovers the spectral radius in the stable regime ($\rho < 1$). Close to the transition from stable to unstable ($\rho = 1$, black contour line), recovery of the spectral radius is approximately achieved (Fig. 4.11 N).

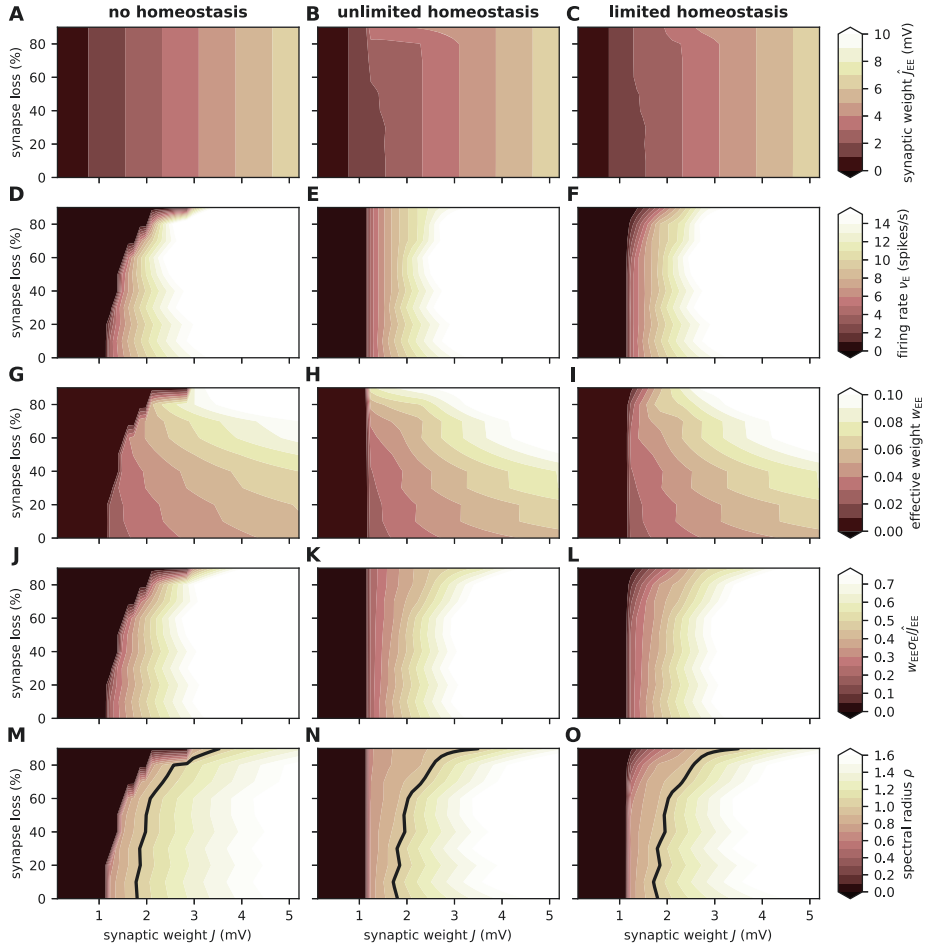


**Figure 4.11: Mean-field theory applied to network with unspecific synapse loss and unspecific synaptic upscaling.** Dependence of the synaptic weight $\hat{J}_{EE}$ (**A–C**), the average firing rate $\nu_E$ of the excitatory population (**D–F**), the effective weight $w_{EE}$ of EE connections (**G–I**), the ratio $w_{EE}\sigma_E/\hat{J}_{EE}$ (**J–L**), and the spectral radius $\rho$ (**M–O**) on the synaptic weight $J$ and the degree of synapse loss in the absence of homeostatic compensation (left column), as well as with unlimited (middle column) and limited firing rate homeostasis (right column). Superimposed black curves in (M–O) mark instability lines $\rho = 1$. Same parameters as in network simulations (see Sec. 4.5.1 and Sec. 4.5.2).

# Chapter 5

# Discussion

## 5.1 Summary of the results

As discussed in the introduction of this thesis, the interaction between rich homeostasis and the pathology of a disease, may result in processes that mask the disease progression and can, consequently, hamper its understanding and diagnosis. I have argued that the involvement of the homeostatic system might also be an important reason for the limited understanding in Alzheimer's disease (AD). In the Chs.2-4, I exemplify how these problems can be tackled from different perspectives and brain scales. In this section, I will summarize the main findings of these chapters and explain how these can contribute to further our understanding of AD, particularly with respect to homeostasis.

A primary problem associated with a rich disease compensation is the observation that marginal distributions of single disease-associated variables often overlap significantly across health conditions (see Sec.1.5). As a consequence, the analysis of the joint distribution of multiple variables is a prerequisite for successful disease diagnosis and understanding. How an approximation of such a joint distribution can be derived is explained in Ch.2. Here, I present the entire pipeline of having a multidimensional data set, constructing and comparing multidimensional statistical models, calculating the probability that a patient has a certain health condition (even considering the combination with other diagnostic criteria) and, ultimately how this approach can aid the decision making process to reach the best possible treatment.

How the exact realization of all these steps look like, is demonstrated based on functional connectivity derived from fMRI data (see Sec.1.4.1). In this chapter, as an exception, the data is not taken from AD patients, but from individuals suffering from schizophrenia, which have been proven to feature functionally disconnected brain areas and could thus provide an important baseline to test the methods employed. The special feature of this chapter is that the likelihoods (see Sec.2.1) are derived step-by-step from first principles through a sequence of easily understandable assumptions of partial exchangeability, sufficient statistics

and prior knowledge. This allows the analysis to be easily extendable to other types of data in a straightforward manner.

Apart from disease diagnosis, the construction of multidimensional statistical models can also help to improve disease understanding. For instance, the role of individual variables, each adding a dimension to the model, can be investigated and scrutinized. If, for example, the removal of such a variable (which reduces the dimensionality of the model), heavily aggravates the separability of the considered health condition, this variable might be particularly affected by the disease, and may thus indicate some of its most critical features. In addition, multidimensional statistical models might help to identify disease subtypes, which may be characterized by regions in this multidimensional space in which the probability of having a specific disease is high and the probability of being healthy is low. Different subtypes would correspond to different regions that only marginally overlap.

After having established a framework for constructing and comparing different statistical models in Ch.2, the resulting knowledge is subsequently applied in Ch.3. The goal of this work has been to evaluate and compare potential frameworks that might be helpful for AD diagnosis. To be more explicit, fMRI data of individuals corresponding to one of tree health conditions AD, MCI or healthy controls (see Sec.1.1), are used to construct cortex–wide connectivity graphs. In these brain graphs, nodes correspond to brain areas and edges describe the functional connectivity between these areas. These graph construction steps can be executed with different methods and parameter configurations. Such graphs are thus constructed in different ways by systematically combining a range of different methods. After graph construction, graph structures are characterized based on different properties, such as shortest path, clustering coefficient etc. (see Sec.1.4.2). Thus, for each combination of graph construction, I obtain a range of graph properties, which build the basis for further analysis. I subsequently investigate which graph construction methods bring about graph properties that are best suited for AD diagnosis. In addition, the informative value of significant differences in graph properties across health conditions is studied.

The analysis of the particular data set investigated in Ch.3 suggests that graphs in which nodes represent large brain areas and edges are described by either correlations or by mutual information transfer yield best diagnostic power. Slight thresholding of the edge weights has only minimal effects on diagnostic performance. Moreover, significant differences in graph properties are not necessarily a good predictor for diagnostic capability.

Despite the evaluation of the different graph construction methods with respect to disease diagnosis, a particular finding in this chapter is that significantly contradicting results of graph properties across health conditions can be a consequence of different graph construction techniques. Thus, depending on the definition of graph nodes and edges, different relation among graph properties are found, e.g. shorter and larger average shortest paths for AD patients. This observation provides a possible explanation for the observed divergences across other studies, in which, for example, the characteristic graph length or clustering coefficient have been reported to be significantly larger (Zhao et al., 2012), unchanged (Sanz-Arigita et al., 2010; Supekar et al., 2008) or significantly smaller (Supekar et al., 2008; Sanz-Arigita et al., 2010) in AD. In all these studies, different graph construction methods have been

applied, but almost no validation on how the graph construction methods could influence the results has taken place.

Consequently, biologically mechanistic interpretations can only be drawn if they include an explanation about the influence of the particular graph construction method on the study outcome. However, due to the emergent nature of the graph properties in graphs derived from fMRI data, this is hardly possible and unless more insight in how for example the exact brain clustering influences graph properties is achieved, such an interpretation should be avoided.

In addition, the sensitivity of graph properties and their relations with the graph construction methods raises doubts that a homeostatic regulation of whole brain functional network exists at all. If it exists, how does the brain define its 'regions' and 'functional connectivity'? Or is it possible to have a global homeostatic regulation that does not require the parcellation of the brain, such that the concept of brain graphs is not appropriate?

In Ch.4, the effects of AD and a specific type of homeostasis on computational characteristics are investigated by means of simple, neuronal network models. AD has been incorporated by deleting excitatory-excitatory synapses. This type of synapse loss reduces the network's firing rate. Referring to observations of homeostatic mechanisms in the human brain, remaining excitatory-excitatory synapses are increased in order to retain the firing rate. With respect to functional characteristics, this kind of AD realization is shown to increase the regularity and to decrease the synchrony in population firing activity. These phenomena are accompanied by a decreased sensitivity to perturbation (increased stability). All these dynamical characteristics recover for a wide range of synapse loss, if homeostatic regulation is not limited. However, if synapse loss is limited and the target firing rate cannot be achieved anymore, dynamical characteristics resemble the characteristics of an AD network without homeostasis. These observation are supported and elucidated by an analogous analytical model and hint towards a successful compensation of AD pathology, if the damage is still small. Thus, if the brain's intrinsic homeostatic strategies are able to slow down AD pathology, finding ways to support homeostasis can constitute an important strategy to slow disease progression and achieve a successful AD treatment.

## 5.2 The reciprocal interactions of different brain scales in Alzheimer's disease

Research that focuses on finding useful therapeutic targets against AD mainly investigates molecular interactions and focuses on finding an appropriate molecular substrate to which a potential drug can bind to. However, only considering processes on very small scales might be too one-sided, as I will demonstrate based on the two following examples.

The interaction of a potential drug with the molecular machinery is mainly investigated in mouse models, which only partly account for the AD pathology observed in humans as AD does not occur in most animal models naturally and has to be induced artificially (see Sec.1.1).

In humans, it is not easy to directly infer the effect of a drug on molecular level. This is because an investigation of molecular interactions with high temporal and spatial resolution relies on methods that are either massively invasive or require the killing of the animal.

In humans, methods that measure diffuse molecular concentrations on a large spacial scale are mainly lumbar puncture and positron-emission-tomography (PET). In addition laborious post-mortem investigations are possible in order to infer local molecular concentrations at the time the patient died. Lumbar puncture is used in order to measure the concentrations of various molecules (e.g. A$\beta$, tau), which are altered in the cerebrospinal fluid (CSF) in AD patients and which represent, to some extent, altered molecular concentrations in the brain (Blennow & Hampel, 2003). However, some proteins involved in AD pathology cannot be investigated via lumbar puncture. Possible reasons are that the concentrations of these molecules in the CSF are unchanged, even if altered in the brain tissue or that the molecules are not stable enough in order to extract them from the CSF (Trombetta et al., 2018). PET imaging, on the other hand, enables the monitoring of glucose, A$\beta$ and tau with rater low spacial resolutions (Marcus et al., 2014). Additionally, it might be possible to extend the repertoire to other molecules involved in AD pathology, provided adequate radioactive markers can be introduced.

In addition to measuring molecular concentrations, researchers have the possibility to non-invasively image structural alterations (e.g. structural MRI) and functional changes (e.g. functional MRI, EEG, MEG) of the brain on various temporal and spacial resolutions. However, the relationship between small-scale, molecular alterations and their effects on large scale structural and functional changes is still difficult to infer. As a consequence, it is difficult to extrapolate the results obtained in pharmacological studies obtained from animal models to human beings. Therefore, once a drug is applied to humans, its effects are primarily evaluated with respect to a potential amelioration of the cognitive decline observed in AD patient and eventually, with respect to brain imaging. If a cognitive improvement is observed and brain images of the treated patient are very similar to the one of the healthy patient, it is not farfetched to conclude that the drug has a positive influence in reversing the changes caused by AD and brings the brain back to normal function. If cognition is ameliorated and the corresponding images are dissimilar, it might also be that the drug causes the brain to move to a state which is different from the healthy state but is still associated with cognitive improvement. In order to close this gap in understanding, it is important to intensify research that focuses on understanding the interaction between the different scales.

Another reason why the focus in AD medical research should be broadened towards bridging different scales is provided by the observation that cognitive performance can be improved in AD patients by means of physical activity and cognitive training (Chen et al., 2016; Choi & Twamley, 2013; Venturelli et al., 2011). There is compelling evidence that physical exercise improves cognitive fitness even more than the drugs that are so far available (Smith et al., 2013). Physical activity has a positive influence on many aspects that have been shown to be affected in AD: it reduces oxidative stress, boosts energy metabolism, increases vascularization and fosters neurogenesis (Choi & Twamley, 2013). Cognitive training and physical exercise entail a change in the activity of various brain areas, which goes hand-in-hand with changes on the molecular levels (Smith et al., 2013; Cespón et al., 2018). Thus, to a certain extent, AD pathology can be counteracted by a top-down treatment with relatively

few side-effects and an improvement of whole-body fitness. Since changing the pattern of (global) brain activity via cognitive and physical training improves cognitive performance, knowledge about the molecular mechanisms involved can even inspire future drug discovery.

In this thesis, I found that alterations in the performance and the dynamics of a small network due to EE-synapses loss (AD) can be compensated by firing rate homeostasis based on increasing the strength of remaining EE-synapses. Can the fMRI scan of an AD brain, in which homeostasis counteracts synapse loss, still look different from a healthy one? As described in Sec.1.4.1, the BOLD signal changes with neuronal activity, which requires a lot of energy. Most of the energy is required for synaptic transmission (for ion pumps), especially of excitatory synapses (Harris et al., 2012). To my knowledge it is not known whether small but many synapses ('healthy state') have different energy requirements than few but large synapses ('AD with homeostasis'). But if this is the case, I would expect that the fMRI signal of the corresponding brain areas of the homeostatic brain is different from the healthy brain. As a consequence, the fMRI image could be used to diagnose AD even in the preclinical phase (assuming that homeostatic compensation successfully delays the onset of cognitive impairment).

So far, I have only considered the damage of brain areas itself and its relation to functional changes as those observed with fMR imaging. I have not discussed what happens if long range connections (the connection between areas) are disrupted, which seems to occur already in early disease stages (Shao et al., 2012). Most of the long range connections are connections between excitatory Pyramidal neurons (McGuire et al., 1991). So, what happens to the dynamics and functionality of such a large-scale network if these excitatory-excitatory connections fade away. Some hints to that questions are provided in Ch.4. So far, I claimed that the results in Ch.4 only account for meso-scale networks of hundreds up to thousands of neurons. I show that the observation of the simulation of the spiking network qualitatively matches the outcome of the theory, which describes the activity of single 'neurons' as a continuous signal. This continuous signal could also describe the activity of brain areas, and the excitatory-excitatory synapses loss, the decay of long range connection observed. The adequate compensation would be the increase of excitatory long range connections, which has not been observed. In our observation the AD network becomes less sensitive to perturbation. Transferred to the large-scale network, this means that small changes in the input to an area do not result in a change of its dynamics. In addition, the perturbation does not propagate to other brain areas. This holding on to a specific network dynamic might be linked to the repetitive behavior (Cullen et al., 2005) observed in AD patients. Accordingly, external influences that would normally stop the repetitive behavior, fade away and do not propagate to the network.

## 5.3 Outlook

Throughout this thesis, I have investigated AD by means of the analysis of fMRI data and through a neuronal network model. In both cases, I have encountered problems, which, in my opinion, represent characteristic shortcomings in AD research. In this section, I will

outline these problems and explain why they are so crucial and how they might be addressed in future work.

With respect to the analysis of fMRI data, I have faced the problem of small data sets, both in terms of numbers of patients (data points) and recording length. Additionally, these datasets typically originate from only a single scanner, which may introduce confounding effects. The problem of small data sets is very common and probably a major source for contradicting research results. First solutions for this problem are provided by multi–site studies that take their data from different imaging modalities available to all qualified researchers for free (e.g. Alzheimer's Disease Neuroimaging Initiative, ADNI [1]). However, there are still too few studies that volunteer to share their data, which slows down research and increases research costs. The too short recording time of the fMRI scans occurs because an individual can spend only a limited time in the fMRI scanner. The short scan period are mainly a problem because the signal-to-noise ratio of the fMRI scan is low and at least 350 sample points have been proposed to be minimal for a successful data analysis assuming a signal-to-noise ratio of 50 (Murphy et al., 2007). This would often require doubling the recording times that are commonly employed. However, since the sampling rates of the scanner get faster, more sampling points might be reached soon without prolonging recording times.

I have experienced another, more general drawback in the analysis of fMRI data - the lack of standard protocols for fMRI data preprocessing. This mainly includes spacial and temporal filtering, noise reduction and registration procedures. Although softwares such as FSL [2] or SPM [3] provide nice graphical user interfaces, which facilitate data analysis, it is often not clear, what kind of parameters (e.g. the standard deviation of a Gaussian filter) should be used for what particular type of data. A general recommendation of how data should be preprocessed would also simplify the comparison between different studies. Hence, applying standard data preprocessing across different studies would rule out the possibility that different results originate from different preprocessing procedures.

Apart from problems with the analysis of fMRI data, I also encountered shortcoming in AD research, when I designed the numerical and analytical model of AD (Ch.4). Here, problems mainly arose because only little is know about the exact process of synaptic decay in AD. Thus, it is still under debate, which synapse type is affected in which brain area at which time point in the course of the disease. This criticism can be extended towards other network and neuron properties such as network firing rates, number of death neurons etc. In my opinion, this gap in knowledge occurs because too many studies focus on animal models of limited reliability and post-mortem brain investigations of human brains are sparse. As a consequence, research results from animal models or cell cultures are often generalized to humans, which might not be justified. A first step to tackle this problem would be a systematic classification of the results with respect to the exact animal model, cell culture conditions, post-mortem analysis etc. Afterwards, similarities and differences across the different classes can be identified. This analysis would facilitate the understanding of which results are rather model specific or which observations apply to all models in the same way.

[1] http://adni.loni.usc.edu/
[2] https://fsl.fmrib.ox.ac.uk/fsl/fslwiki
[3] https://www.fil.ion.ucl.ac.uk/spm/

This understanding is an important requirement in order to estimate in how far results from animal models and cell cultures can be transferred to humans.

So far, I have explained the problems in AD research that have been directly linked to the studies of this thesis. At the end, I would like to raise a more general problem in AD research. As already mentioned earlier, the disease incidence can be influenced by a range of lifestyle factors, e.g. smoking, physical activity, leisure activity, education (for review see Mayeux & Stern, 2012). But this knowledge is difficult to transfer to society (Pope et al., 2003). In general, I have the impression that people are more interested in taking a drug, rather than changing their lifestyle. Personally, I am in favor of projects that give elderly people more responsibilities and challenges especially with respect to childcare. The reasons are the following: it is argued that AD is the price humans pay for having a very long post-fertile life span (Gunn-Moore et al., 2018). The purpose of this extended life is thought to take of the task of childcare of grandchildren. Childcare usually implies body movement, cognitive training and dealing with new challenges and social integration. All these factors have been identified to be beneficial for AD treatment (Mayeux & Stern, 2012; Pope et al., 2003; Hsiao et al., 2018), because they probably promote intrinsic compensation mechanisms (Arenaza-Urquijo et al., 2015). In addition, a change in lifestyle may also be beneficial for other diseases and improves life quality in general.

# Bibliography


Abbott, L., DePasquale, B., & Memmesheimer, R.-M. (2016). Building functional networks of spiking model neurons. *Nat Neurosci 19*(3), 350–355.

Abeles, M. (1982). *Local Cortical Circuits: An Electrophysiological Study*. Studies of Brain Function. Berlin, Heidelberg, New York: Springer-Verlag.

Abuhassan, K., Coyle, D., Belatreche, A., & Maguire, L. (2012). Compensating for synaptic loss in Alzheimer's disease. *Journal of Computational Neuroscience 36*(1), 19–37.

Aldous, D. J. (1985). Exchangeability and related topics. In Aldous et al. (1985), pp. VII, 1–198. `https://www.stat.berkeley.edu/~aldous/206-Exch/Papers/papers.html`.

Aldous, D. J., Ibragimov, I. A., & Jacod, J. (1985). *École d'Été de Probabilités de Saint-Flour XIII – 1983*, Volume 1117 of *Lecture notes in mathematics*. Berlin: Springer. Édité par P. L. Hennequin.

Allen, D. M. (1974). The relationship between variable selection and data agumentation and a method for prediction. *Technometrics 16*(1), 125–127.

Amatniek, J. C., Hauser, A. W., DelCastillo Castaneda, C., Jacobs, D. M., Marder, K., Bell, K., Albert, M., Brandt, J., & Stern, Y. (2006). Incidence and predictors of seizures in patients with Alzheimer's disease. *Epilepsia 47*(5), 867–872.

Amoroso, N., La Rocca, M., Bruno, S., Maggipinto, T., Monaco, A., Bellotti, R., & Tangaro, S. (2017). Brain structural connectivity atrophy in Alzheimer's disease. *arXiv 1709.02369 [physics.med-ph]*.

Andersen, E. B. (1970). Sufficiency and exponential families for discrete sample spaces. *J. Am. Stat. Assoc. 65*(331), 1248–1255.

Arenaza-Urquijo, E. M., Wirth, M., & Chételat, G. (2015). Cognitive reserve and lifestyle: moving towards preclinical Alzheimer's disease. *Front Aging Neurosci 7*, 134.

Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., & Gee, J. C. (2011). A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage 54*(3), 2033–44.

Baddeley, A. D., Baddeley, H. A., Bucks, R. S., & Wilcock, G. K. (2001). Attentional control in Alzheimer's disease. *Brain 124*(Pt8), 1492–1508.

Baldwin, S., & Tomaszewski Farias, S. (2009). Unit 10.3: assessment of cognitive impairments in the diagnosis of Alzheimer's disease. *Curr Protoc Neurosci 0 10*, Unit10.3.

Ballatore, C., Lee, V. M.-Y., & Trojanowski, J. Q. (2007). Tau-mediated neurodegeneration in Alzheimer's disease and related disorders. *Nature Reviews Neuroscience 8*, 663–672.

Barak, O., Sussillo, D., Romo, R., Tsodyks, M., & Abbott, L. F. (2013). From fixed points to chaos: three models of delayed discrimination. *Prog. Neurobiol. 103*, 214–222.

Barankin, E. W., & Maitra, A. P. (1963). Generalization of the Fisher-Darmois-Koopman-Pitman theorem on sufficient statistics. *Sankhyā A 25*(3), 217–244.

Barnard, G. A., Atkinson, A. C., Chan, L. K., Dawid, A. P., Downton, F., Dickey, J., Baker, A. G., Barndorff-Nielsen, O., Cox, D. R., Geisser, S., Hinkley, D., Hocking, R. R., Young, A. S., & Stone, M. (1974). Discussion [Cross-Validatory choice and assessment of statistical predictions]. *J. Roy. Stat. Soc. B 36*(2), 133–147. See Stone (1974).

Barnard, J., McCulloch, R., & Meng, X.-L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Stat. Sinica 10*(4), 1281–1311.

Barrett, D. G., Denève, S., & Machens, C. K. (2016). Optimal compensation for neuron loss. *eLife 5*, e12454.

Bartlett, M. S. (1952). The statistical significance of odd bits of information. *Biometrika 39*(3–4), 228–237.

Bassett, D. S., Bullmore, E., Verchinski, B. A., Mattay, V. S., Weinberger, D. R., & Meyer-Lindenberg, A. (2008). Hierarchical organization of human cortical networks in health and Schizophrenia. *J. Neurosci 28*(37), 9239–9248.

Beckmann, C. F., & Smith, S. M. (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Transactions on Medical Imaging 23*(2), 137–152.

Beggs, J. M., & Plenz, D. (2003). Neuronal avalanches in neocortical circuits. *J. Neurosci. 23*(35), 11167–11177.

Belleville, S., Bherer, L., Lepage, E., Chertkow, H., & Gauthier, S. (2008). Task switching capacities in persons with Alzheimer's disease and mild cognitive impairment. *Neuropsychologia 46*(8), 2225–2233.

Bennett, D. A., Wilson, R. S., Schneider, J. A., Evans, D. A., Beckett, L. A., Aggarwal, N. T., Barnes, L. L., Fox, J. H., & Bach, J. (2002). Natural history of mild cognitive impairment in older persons. *Neurology 59*(2), 198–205.

Berger, J. O., & Pericchi, L. R. (1996). The intrinsic Bayes factor for model selection and prediction. *J. Am. Stat. Assoc. 91*(433), 109–122.

Berk, R. H. (1966). Limiting behavior of posterior distributions when the model is incorrect. *Ann. Math. Stat. 37*(1), 51–58.

Bernardo, J.-M. (1979). Expected information as expected utility. *Ann. Stat. 7*(3), 686–690. http://www.uv.es/~bernardo/publications.html.

Bernardo, J.-M., DeGroot, M. H., Lindley, D. V., & Smith, A. F. M. (Eds.) (1988). *Bayesian Statistics 3*. Oxford: Oxford University Press.

Bernardo, J.-M., & Smith, A. F. (2000). *Bayesian Theory* (reprint ed.). Wiley series in probability and mathematical statistics. New York: Wiley. First publ. 1994.

Bertschinger, N., & Natschläger, T. (2004). Real-time computation at the edge of chaos in recurrent neural networks. *Neural Comput. 16*(7), 1413–1436.

Besag, J., Bickel, P. J., Brøns, H., Fraser, D. A. S., Reid, N., Helland, I. S., Huber, P. J., Kalman, R., Pincus, S., Tjur, T., & McCullagh, P. (2002). What is a statistical model?: Discussion and Rejoinder. *Ann. Stat. 30*(5), 1267–1310. http://www.stat.uchicago.edu/~pmcc/publications.html. See McCullagh (2002).

Binzegger, T., Douglas, R. J., & Martin, K. A. C. (2004). A quantitative map of the circuit of cat primary visual cortex. *J. Neurosci. 24*(39), 8441–8453.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Information science and statistics. New York: Springer.

Blennow, K., & Hampel, H. (2003). CSF markers for incipient Alzheimer's disease. *The Lancet Neurology 2*(10), 605–613.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics 10*, P10008.

Boerlin, M., Machens, C. K., & Denève, S. (2013). Predictive coding of dynamical variables in balanced spiking networks. *PLoS Computational Biology 9*(11), e1003258.

Bordier, C., Nicolini, C., & Bifone, A. (2017). Graph analysis and modularity of brain functional connectivity networks: searching for the optimal threshold. *Frontiers in Neuroscience 11*, 441.

Boucsein, C., Tetzlaff, T., Meier, R., Aertsen, A., & Naundorf, B. (2009). Dynamical response properties of neocortical neuron ensembles: multiplicative versus additive noise. *J. Neurosci. 29*(4), 1006–1010.

Boyd, S., & Chua, L. (1985). Fading memory and the problem of approximating nonlinear operators with volterra series. *IEEE Transactions on Circuits and Systems 32*, 1150–1165.

Browder, F. E. (Ed.) (1992). *Mathematics into the Twenty-first Century: 1988 Centennial Symposium August 8–12*, Volume II of *American Mathematical Society centennial publications*. Providence, USA: American Mathematical Society.

Brun, A., Liu, X., & C., E. (1995). Synapse loss and gliosis in the molecular layer of the cerebral cortex in Alzheimer's disease and in frontal lobe degeneration. *Neurodegeneration 4*(2), 171–177.

Brunel, N. (2000). Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *J. Comput. Neurosci. 8*(3), 183–208.

Brunel, N., & Hakim, V. (1999). Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural Comput. 11*(7), 1621–1671.

Bruno, A. (1964). On the notion of partial exchangeability. *Giorn. Ist. Ital. Att. 27*, 174–196. Transl. in de Finetti (1972), Ch. 10, pp. 229–246.

Buesing, L., Bill, J., Nessler, B., & Maass, W. (2011). Neural dynamics as sampling: a model for stochastic computation in recurrent networks of spiking neurons. *PLoS Comp Biol 7*(11), e1002211.

Busche, M. A., Chen, X., Henning, H. A., Reichwald, J., Staufenbiel, M., Sakmann, B., & Konnerth, A. (2012). Critical role of soluble amyloid-$\beta$ for early hippocampal hyperactivity in a mouse model of Alzheimer's disease. *Proceedings of the National Academy of Sciences 109*(22), 8740–5.

Buxton, R. B. (2013). The physics of functional magnetic resonance imaging (fMRI). *Rep Prog Phys 76*(9), :096601.

Cabanac, M. (2006). Adjustable set point: to honor Harold T. Hammel. *Journal of Applied Physiology 100*(4), 1338–1346.

Canter, R. G., Penney, J., & Tsai, L.-H. (2016). The road to restoring neural circuits for the treatment of Alzheimer's disease. *Nature 539*, 187–196.

Çetin, M. S., Christensen, F., Abbott, C. C., Stephen, J. M., Mayer, A. R., Cañive, J. M., Bustillo, J. R., Pearlson, G. D., & Calhoun, V. D. (2014). Thalamus and posterior temporal lobe show greater inter-network connectivity at rest and across sensory paradigms in Schizophrenia. *NeuroImage 97*, 117–126.

Çetin, M. S., Houck, J. M., Rashid, B., Agacoglu, O., Stephen, J. M., Sui, J., Canive, J., Mayer, A., Aine, C., Bustillo, J. R., & Calhoun, V. D. (2016). Multimodal classification of Schizophrenia patients with MEG and fMRI data using static and dynamic connectivity measures. *Frontiers in Neuroscience 10*, 466.

Cespón, J., Miniussi, C., & Pellicciari, M. C. (2018). Interventional programmes to improve cognition during healthy and pathological ageing: cortical modulations and evidence for brain plasticity. *Ageing Research Reviews 43*, 81–98.

Chakravarthy, N., Tsakalis, K., Sabesan, S., & Iasemidis, L. (2009). Homeostasis of brain dynamics in epilepsy: a feedback control systems perspective of seizures. *Annals of Biomedical Engineering 37*(3), 565–568.

Chen, M., Mecca, A., Naganawa, M., Finnema, S., Toyonaga, T., Lin, S., Najafzadeh, S., Ropchan, J., Lu, Y., McDonald, J., Michalak, H., Nabulsi, N., Arnsten, A., Huang, Y., Carson, R., & van Dyck, C. (2018). Assessing synaptic density in Alzheimer disease with synaptic vesicle glycoprotein 2a positron emission tomographic imaging. *JAMA Neurology 75*, 1215–1224.

Chen, W.-W., Zhang, X., & Huang, W.-J. (2016). Role of physical exercise in Alzheimer's disease. *Biomedical Reports 4*(4), 403–407.

Chertkow, H., Feldman, H. H., Jacova, C., & Massoud, F. (2013). Definitions of dementia and predementia states in Alzheimer's disease and vascular cognitive impairment: consensus from the Canadian conference on diagnosis of dementia. *Alzheimer's Research & Therapy 5*(1), S2.

Chicharro, D. (2011). On the spectral formulation of Granger causality. *Biol. Cybern. 105*(5–6), 331–347.

Choi, J., & Twamley, E. W. (2013). Cognitive rehabilitation therapies for Alzheimer's disease: a review of methods to improve treatment engagement and self–efficacy. *Neuropsychol. Rev 23*(1), 48–62.

Cifarelli, D. M., & Regazzini, E. (1982). Some considerations about mathematical statistics teaching methodology suggested by the concept of exchangeability. In Koch & Spizzichino (1982), pp. 185–205.

Collins, D. L., Holmes, C. J., Peters, T. M., & Evans, A. C. (1995). Automatic 3-D model-based neuroanatomical segmentation. *Human Brain Mapping 3*(3), 190–208.

Connor, J. T., Martin, R. D., & Atlas, L. E. (1994). Recurrent neural networks and robust time series prediction. *Trans. Neur. Netw. 5*(2), 240–254.

Corbett, B. F., Leiser, S. C., Ling, H.-P., Nagy, R., Breysse, N., Zhang, X., Hazra, A., Brown, J. T., Randall, A. D., Wood, A., Pangalos, M. N., Reinhart, P. H., & Chin, J. (2013). Sodium channel cleavage is associated with aberrant neuronal activity and cognitive deficits in a mouse model of Alzheimer's disease. *Journal of Neuroscience 33*, 7020–7026.

Cowan, R. L., & Wilson, C. J. (1994). Spontaneous firing patterns and axonal projections of single corticostriatal neurons in the rat medial agranular cortex. *J. Neurophysiol. 71*, 17–32.

Cox, R. T. (1946). Probability, frequency, and reasonable expectation. *Am. J. Phys. 14*(1), 1–13. `http://algomagic.org/ProbabilityFrequencyReasonableExpectation.pdf`.

Cox, R. T. (1961). *The Algebra of Probable Inference*. Baltimore: The Johns Hopkins Press.

Cox, R. T. (1979). Of inference and inquiry, an essay in inductive logic. In Levine & Tribus (1979), pp. 119–167.

Crane, P. K., Walker, R., Hubbard, R. A., Li, G., Nathan, D. M., Zheng, H., Haneuse, S., Craft, S., Montine, T. J., Kahn, S. E., McCormick, W., McCurry, S. M., Bowen, J. D., & Larson, E. B. (2013). Glucose levels and risk of dementia. *N Engl J Med 369*(6), 540–548.

Cullen, B., Coen, R. F., Lynch, C. A., Cunningham, C. J., Coakley, D., Robertson, I. H., & Lawlor, B. A. (2005). Repetitive behaviour in Alzheimer's disease: description, correlates and functions. *Int. J. Geriat. Psychiatry 20*(7), 686–693.

da Silva, F. L., Pijn, J. P., & Boeijinga, P. (1989). Interdependence of EEG signals: linear vs. nonlinear associations and the significance of time delays and phase shifts. *Brain Topography 2*(1/2), 9–18.

Dagher, N. N., Najafi, A. R., Kayala, K. M. N., Elmore, M. R. P., White, T. E., BMedeiros, R., West, B. L., & West, B. L. (2015). Colony-stimulating factor 1 receptor inhibition prevents microglial plaque association and improves cognition in 3xTg-AD mice. *Journal of Neuroinflammation 12*(1), 139.

Dambre, J., Verstraeten, D., Schrauwen, B., & Massar, S. (2012). Information processing capacity of dynamical systems. *Scientific reports 2*, 514.

Damien, P., Dellaportas, P., Polson, N. G., & Stephens, D. A. (Eds.) (2013). *Bayesian Theory and Applications*. Oxford: Oxford University Press.

Damoiseaux, J. S. (2012). Resting-state fMRI as a biomarker for Alzheimer's disease? *Alzheimer's Research & Therapy 4*(2), 8.

Darmois, G. (1935). Sur les lois de probabilité à estimation exhaustive. *Comptes rendus hebdomadaires des séances de l'Académie des sciences 200*, 1265–1266.

Davidoff, F. (1999). Standing statistics right side up. *Ann. Intern. Med. 130*(12), 1019–1021. `http://www.perfendo.org/docs/bayesprobability/5.3_goodmanannintmed99all.pdf`. See Goodman (1999); Sulmasy et al. (2000).

Dawid, A. P. (1982a). Intersubjective statistical models. In Koch & Spizzichino (1982), pp. 217–232.

Dawid, A. P. (1982b). The well-calibrated Bayesian. *J. Am. Stat. Assoc. 77*(379), 605–610.

Dawid, A. P. (2013). Exchangeability and its ramifications. In Damien et al. (2013), Chapter 2, pp. 19–29.

de Finetti, B. (1937). La prévision : ses lois logiques, ses sources subjectives. *Ann. Inst. Henri Poincaré 7*(1), 1–68. Transl. in Kyburg & Smokler (1980), pp. 53–118, by Henry E. Kyburg, Jr.

de Finetti, B. (1938). Sur la condition d'équivalence partielle. In B. de Finetti, V. Glivenko, & G. Neymann (Eds.), *Colloque consacré à la théorie des probabilités. VI : Conceptions diverses*, Number 739 in Actualités scientifiques et industrielles, pp. 5–18. Paris: Hermann. Transl. in Jeffrey (1980), pp. 193–205, by P. Benacerraf and R. Jeffrey.

de Finetti, B. (1972). *Probability, Induction and Statistics: The art of guessing*. London: Wiley.

de Toledo-Morrell, L., Dickerson, B., Sullivan, M., Spanovic, C., Wilson, R., & Bennett, D. (2000). Hemispheric differences in hippocampal volume predict verbal and spatial memory performance in patients with Alzheimer's disease. *Hippocampus 10*, 136–142.

de Wilde, M. C., Overk, C. R., Sijben, J. W., & Masliah, E. (2016). Meta-analysis of synaptic pathology in Alzheimer's disease reveals selective molecular vesicular machinery vulnerability. *Alzheimers Dement 12*(6), 633–44.

DeFelipe, J., & Fariñas, I. (1992). The pyramidal neuron of the cerebral cortex: morphological and chemical characteristics of the synaptic inputs. *Progress in Neurobiology 39*(6), 563–607.

DeGroot, M. H. (2004). *optimal statistical decisions* (reprint ed.). Wiley classics library. New York: Wiley.

DeKosky, S., & Scheff, S. (1990). Synapse loss in frontal cortex biopsies in Alzheimer's disease: correlation with cognitive severity. *Ann Neurol. 27*, 1531 – 8249.

Demirci, O., Clark, V. P., & Calhoun, V. D. (2008). A projection pursuit algorithm to classify individuals using fMRI data: Application to schizophrenia. *NeuroImage 39*(4), 1774–1782.

Dennis, E. L., & Thompson, P. M. (2014). Functional brain connectivity using fMRI in aging and Alzheimer's disease. *Neuropsychology review 24*(1), 49–62.

Denny, J. L. (1967). Sufficient conditions for a family of probabilities to be exponential. *Proc. Natl. Acad. Sci. (USA) 57*(5), 1184–1187.

Desai, N. S., Rutherford, L. C., & Turrigiano, G. G. (1999). Plasticity in the intrinsic excitability of cortical pyramidal neurons. *Nat. Neurosci. 2*(6), 515–520.

Desikan, R. S., Cabral, H. J., Hess, C. P., Dillon, W. P., Glastonbury, C. M., Weiner, M. W., Schmansky, N. J., Greve, D. N., Salat, D. H., Buckner, R. L., & Fischl, B. (2009). Automated MRI measures identify individuals with mild cognitive impairment and Alzheimer's disease. *Brain 132*(8), 2048–57.

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., & Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage 31*(3), 968–980.

Diaconis, P. (1988). Recent progress on de Finetti's notions of exchangeability. In Bernardo et al. (1988), pp. 111–125. With discussion by D. Blackwell, Simon French, and author's reply. `http://statweb.stanford.edu/~cgates/PERSI/year.html`, `https://statistics.stanford.edu/research/recent-progress-de-finettis-notions-exchangeability`.

Diaconis, P. (1992). Sufficiency as statistical symmetry. In Browder (1992), pp. 15–26. First publ. 1991 as technical report `https://statistics.stanford.edu/research/sufficiency-statistical-symmetry`.

Diaconis, P., & Freedman, D. (1980). De Finetti's generalizations of exchangeability. In Jeffrey (1980), pp. 233–249.

Diaconis, P., & Freedman, D. (1981). Partial exchangeability and sufficiency. In Ghosh & Roy (1981), pp. 205–236. `http://statweb.stanford.edu/~cgates/PERSI/year.html`. Also publ. 1982 as technical report `https://www.stat.berkeley.edu/~aldous/206-Exch/Papers/diaconis_freedman_PES.pdf`.

Diaconis, P., & Freedman, D. (1986). On the consistency of Bayes estimates. *Ann. Stat. 14*(1), 1–26.

Diaconis, P., & Ylvisaker, D. (1979). Conjugate priors for exponential families. *Ann. Stat. 7*(2), 269–281.

Dickerson, B., Salat, D., Greve, D., Chua, E., E, R.-G., Rentz, D., Bertram, L., Mullin, K., R, T., Blacker, D., Albert, M., & Sperling, R. (2005). Increased hippocampal activation in mild cognitive impairment compared to normal aging and AD. *Neurology 65*(3), 404–11.

Dillen, K. N., Jacobs, H. I., Kukolja, J., Richter, N., von Reutern, B., Onur, O. A., Langen, K.-J., & Fink, G. R. (2017). Functional disintegration of the default mode network in prodromal Alzheimer's disease. *Journal of Alzheimer's Disease 59*(1), 169–187.

Dorostkar, M. M., Zou, C., Blazquez-Llorca, L., & Herms, J. (2015). Analyzing dendritic spine pathology in Alzheimer's disease: problems and opportunities. *Acta Neuropathol 130*(1), 1–10.

Dupré, M. J., & Tipler, F. J. (2009). New axioms for rigorous Bayesian probability. *Bayesian Anal. 4*(3), 599–606.

Ecker, A. S., Berens, P., Keliris, G. A., Bethge, M., Logothetis, N. K., & Tolias, A. S. (2010). Decorrelated neuronal firing in cortical microcircuits. *Science 327*(5965), 584–587.

Edgeworth, F. Y. (1898). On the representation of statistics by mathematical formulæ (Part I). *J. Roy. Stat. Soc. 61*(4), 670–700. See also Edgeworth (1899, 1900).

Edgeworth, F. Y. (1899). On the representation of statistics by mathematical formulæ (Part II & III). *J. Roy. Stat. Soc. 62*(1, 2), 125–140, 373–385. See also Edgeworth (1898, 1900).

Edgeworth, F. Y. (1900). On the representation of statistics by mathematical formulæ (Supplement). *J. Roy. Stat. Soc. 63*(1), 72–81. See also Edgeworth (1898, 1899).

Eliasmith, C., & Anderson, C. H. (2004). *Neural engineering: Computation, representation, and dynamics in neurobiological systems*. MIT press.

Ellison-Wright, I., & Bullmore, E. (2009). Meta-analysis of diffusion tensor imaging studies in Schizophrenia. *Schizophrenia Research 108*(1), 3–10.

Ellison-Wright, I., & Bullmore, E. (2010). Anatomy of bipolar disorder and Schizophrenia: a meta-analysis. *Schizophrenia Research 117*(1), 1–12.

Engelken, R., Farkhooi, F., Hansel, D., van Vreeswijk, C., & Wolf, F. (2015). Comment on "two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons". *bioRxiv*, 017798.

Eslamizade, M., Saffarzadeh, F., Mousavi, S., Meftahi, G., Hosseinmardi, N., Mehdizadeh, M., & Janahmadi, M. (2015). Alterations in CA1 pyramidal neuronal intrinsic excitability mediated by Ih channel currents in a rat model of amyloid beta pathology. *Neuroscience 305*, 279–292.

Ferguson, T. S. (1974). Prior distributions on spaces of probability measures. *Ann. Stat. 2*(4), 615–629.

Fernandes, D., & Luísa, C. A. (2016). Mechanisms of homeostatic plasticity in the excitatory synapse. *J. Neurochem. 139*(6), 973–996.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Phil. Trans. R. Soc. Lond. A 222*, 309–368. `http://www.stats.org.uk/statistical-inference/Fisher1922.pdf`.

Fourcaud, N., & Brunel, N. (2002). Dynamics of the firing probability of noisy integrate-and-fire neurons. *Neural Comput. 14*(9), 2057–2110.

Fraser, D. A. S. (1963). On sufficiency and the exponential family. *J. Roy. Stat. Soc. B 25*(1), 115–123.

Freedman, D. A. (1962). Invariants under mixing which generalize de Finetti's theorem. *Ann. Math. Stat. 33*(3), 916–923.

Frere, S., & Slutsky, I. (2017). Alzheimer's disease: from firing instability to homeostasis network collapse. *Neuron 97*(1), 32–58.

Friedman, N., Ito, S., Brinkman, B. A. W., Shimono, M., DeVille, R. E. L., Dahmen, K. A., Beggs, J. M., & Butler, T. C. (2012). Universal critical dynamics in high resolution neuronal avalanche data. *Phys. Rev. Lett. 108*(20), 208102.

Friston, K. J., & Frith, C. D. (1995). Schizophrenia: a disconnection syndrome? *Clinical Neuroscience 3*, 89–97.

Fröhlich, F., Bazhenov, M., & Sejnowski, T. J. (2008). Pathological effect of homeostatic synaptic scaling on network dynamics in diseases of the cortex. *J Neurosci 28*(7), 1709–1720.

Fu, H. F., Rodriguez, G. A., Herman, M., Emrani, S., Nahmani, E., Barrett, G., Figueroa, H. Y., Goldberg, E., Hussaini, S. A., & Duff, K. E. (2017). Tau pathology induces excitatory neuron loss, grid cell dysfunction, and spatial memory deficits reminiscent of early Alzheimer's disease. *Neuron 93*(3), 533–541.e5.

Furber, S. (2016). Large-scale neuromorphic computing systems. *Journal of Neural Engineering 13*(5), 051001.

Gardner, M. W., & Dorling, S. R. (1998). Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric Environment 32*(14), 2627–1636.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian Data Analysis* (3 ed.). Texts in statistical science. Boca Raton, USA: Chapman & Hall/CRC. First publ. 1995.

Gheiratmand, M., Rish, I., Cecchi, G. A., Brown, M. R. G., Greiner, R., Polosecki, P. I., Bashivan, P., Greenshaw, A. J., Ramasubbu, R., & Dursun, S. M. (2017). Learning stable and predictive network-based patterns of schizophrenia and its clinical symptoms. *NPJ Schizophr. 3*, 22.

Ghosh, J. K., & Roy, J. (Eds.) (1981). *Statistics: Applications and New Directions*. Calcutta: Indian Statistical Institute.

Gibson, J. R., Bartley, A. F., & Huber, K. M. (2006). Role for the subthreshold currents ILeak and IH in the homeostatic control of excitability in neocortical somatostatin-positive inhibitory neurons. *Journal of Neurophysiology 96*(1), 420–432.

Gits, H. C. (2016). Relating connectivity and graph analysis to cognitive function in Alzheimer's disease. *Michigan Journal of Medicine 1*(1).

Godwin, D., Barry, R. L., & Marois, R. (2015). Breakdown of the brain's functional network modularity with awareness. *Proc Natl Acad Sci USA 112*(12), 3799–3804.

Golovko, V., Savitsky, Y., & Manikov, N. (2002). Neural networks for signal processing in measurement analysis and industrial applications : the case of chaotic signal processing.

Good, I. J. (1950). *Probability and the Weighing of Evidence*. London: Griffin.

Good, I. J. (1956). The surprise index for the multivariate normal distribution. *Ann. Math. Stat. 27*(4), 1130–1135. See also corrections Good (1957b).

Good, I. J. (1957a). The appropriate mathematical tools for describing and measuring uncertainty. In Good (1983), Chapter 16, pp. 173–177. First publ. 1957.

Good, I. J. (1957b). Corrections to "The surprise index for the multivariate normal distribution". *Ann. Math. Stat. 28*(4), 1055. See Good (1956).

Good, I. J. (1983). *Good Thinking: The Foundations of Probability and Its Applications*. Minneapolis, USA: University of Minnesota Press.

Goodman, S. N. (1999). Toward evidence-based medical statistics. 1: The $P$ value fallacy. 2: The Bayes factor. *Ann. Intern. Med. 130*(12), 995–1013. `http://www.perfendo.org/docs/bayesprobability/5.3_goodmanannintmed99all.pdf`. See also comments and reply Davidoff (1999); Sulmasy et al. (2000).

Grabner, G., Janke, A. L., Budge, M. M., Smith, D., Pruessner, J., & Collins, D. L. (2006). Symmetric atlasing and model based segmentation: an application to the hippocampus in older adults. *Med Image Comput Comput Assist Interv 9*(Pt2), 58–66.

Grassberger, P., Schreiber, T., & Schaffrath, C. (1991). Nonlinear time sequence analysis. *Int. J. Bifurcation Chaos 1*, 521–547.

Grathwohl, S. A., Kälin, R. E., Bolmont, T., Prokop, S., Winkelmann, G., Kaeser, S. A., Odenthal, J., Radde, R., Eldh, T., Gandy, S., Aguzzi, A., Staufenbiel, M., Mathews, P. M., Wolburg, H., Heppner, F. L., & Jucker, M. (2009). Formation and maintenance of Alzheimer's disease $\beta$ amyloid plaques in the absence of microglia. *Nature Neuroscience 12*, 1361–1363.

Greenland, S. (1998). Probability logic and probabilistic induction. *Epidemiology 9*(3), 322–332. See also Maclure (1998).

Grubb, M. S., & Burrone, J. (2010). Activity-dependent relocation of the axon initial segment fine-tunes neuronal excitability. *Nature 465*(7301), 1070–1074.

Grytskyy, D., Tetzlaff, T., Diesmann, M., & Helias, M. (2013). Invariance of covariances arises out of noise. *AIP Conf. Proc. 1510*, 258–262.

Gräff, J., Rei, D., Guan, J.-S., Wang, W.-Y., Seo, J., Hennig, K. M., Nieland, T. J. F., Fass, D. M., Kao, P. F., Kahn, M., Su, S. C., Samiei, A., Joseph, N., Haggarty, S. J., Delalle, I., & Tsai, L.-H. (2012). An epigenetic blockade of cognitive functions in the neurodegenerating brain. *Nature 483*, 222–226.

Guimerá, R., & Nunes Amaral, L. A. (2005). Functional cartography of complex metabolic networks. *Nature 433*, 895–900.

Gulyás, A. I., Megúas, M., Emri, Z., & Freund, T. F. (1999). Total number and ratio of excitatory and inhibitory synapses converging onto single interneurons of different types in the CA1 area of the rat hippocampus. *J. Neurosci. 19*(22), 10082–10097.

Gunn-Moore, D., Kaidanovich-Beilin, O., Gallego Iradi, M. C., Gunn-Moore, F., & Lovestone, S. (2018). Alzheimer's disease in humans and other animals: a consequence of postreproductive life span and longevity rather than aging. *Alzheimer's & Dementia 14*(2), 195–204.

Gupta, A. K., & Nagar, D. K. (2000). *Matrix Variate Distributions*, Volume 104 of *Monographs and surveys in pure and applied mathematics*. Boca Raton, USA: Chapman & Hall/CRC.

Haghani, M., Janahmadi, M., & Shabani, M. (2012). Protective effect of cannabinoid CB1 receptor activation against altered intrinsic repetitive firing properties induced by A$\beta$ neurotoxicity. *Neuroscience Letters 507*(1), 33–37.

Hailperin, T. (1996). *Sentential Probability Logic: Origins, Development, Current Status, and Technical Applications*. London: Associated University Presses.

Hall, A. M., Throesch, B. T., Buckingham, S. C., Markwardt, S. J., Peng, Y., Wang, Q., Hoffman, D. A., & Roberson, E. D. (2015). Tau-dependent kv4.2 depletion and dendritic hyperexcitability in a mouse model of Alzheimer's disease. *J Neurosci 35*(15), 6221–30.

Halpern, J. Y. (1999). Cox's theorem revisited. *J. Artif. Intell. Res. 11*, 429–435. See also Snow (1998).

Hanuschkin, A., Kunkel, S., Helias, M., Morrison, A., & Diesmann, M. (2010). A general and efficient method for incorporating precise spike times in globally time-driven simulations. *Front. Neuroinform. 4*, 113.

Harish, O., & Hansel, D. (2015). Asynchronous rate chaos in spiking neuronal circuits. *PLoS Comput Biol 11*(7), e1004266.

Harris, J. J., Jolivet, R., & Attwell, D. (2012). Synaptic energy use and supply. *Neuron 75*(5), 762–777.

Harris, K. D., & Thiele, A. (2011). Cortical state and attention. *Nat. Rev. Neurosci. 12*, 509–523.

Hayasaka, S. (2013). Functional connectivity networks with and without global signal correction. *Front. Hum. Neurosci. 7*, :880.

He, Y., Wang, J., Wang, L., Chen, Z. J., Yan, C., Yang, H., Tang, H., Zhu, C., Gong, Q., Zang, Y., & Evans, A. C. (2009). Uncovering intrinsic modular organization of spontaneous brain activity in humans. *PloS One 4*(4), 1–18.

Hedskog, L., Pinho, C. M., Filadi, R., Rönnbäck, A., Hertwig, L., Wiehager, B., Larssen, P., Gellhaar, S., Sandebring, A., Westerlund, M., Graff, C., Winblad, B., Galter, D., Behbahani, H., Pizzo, P., Glaser, E., & Ankarcrona, M. (2013). Modulation of the endoplasmic reticulum–mitochondria interface in Alzheimer's disease and related models. *Proc Natl Acad Sci U S A 110*(19), 7916–7921.

Heiberg, T., Kriener, B., Tetzlaff, T., Casti, A., Einevoll, G., & Plesser, H. (2013). Firing-rate models capture essential response dynamics of LGN relay cells. *J. Comput. Neurosci. 35*(3), 359–375.

Heiberg, T., Kriener, B., Tetzlaff, T., Einevoll, G. T., & Plesser, H. E. (2018). Firing-rate models for neurons with a broad repertoire of spiking behaviors. *J. Comput. Neurosci.* (in press).

Helias, M., Tetzlaff, T., & Diesmann, M. (2013). Echoes in correlated neural systems. *New J. Phys. 15*, 023002.

Herholz, K. (2010). Cerebral glucose metabolism in preclinical and prodromal Alzheimer's disease. *Expert Review of Neurotherapeutics 10*(11), 1667–1673.

Hinton, G., Osindero, S., & Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation 18*(7), 1527–1554.

Hipp, C. (1974). Sufficient statistics and exponential families. *Ann. Stat. 2*(6), 1283–1292.

Ho, B.-C., Andreasen, N. C., Flaum, M., Nopoulos, P., & Miller, D. (2000). Untreated initial psychosis: Its relation to quality of life and symptom remission in first-episode Schizophrenia. *Am. J. Psychiatry 157*(5), 808–815.

Hoenig, M. C., Bischof, G. N., Seemiller, J., Hammes, J., Kukolja, J., Onur, O. A., Jessen, F., Fliessbach, K., Neumaier, B., Fink, G. R., van Eimeren, T., & Drzezga, A. (2018). Networks of tau distribution in Alzheimer's disease. *Brain 141*(2), 568–581.

Hong, S., VF, B.-G., Nfonoyim, B. M., Frouin, A., Li, S., Ramakrishnan, S., Merry, K. M., Shi, Q., Rosenthal, A., Barres, B. A., Lemere, C. A., Selkoe, D. J., & Stevens, B. (2016). Complement and microglia mediate early synapse loss in Alzheimer mouse models. *Science 352*(6286), 712–716.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. (USA) 79*(8), 2554–2558. `http://cns.upf.edu/jclub/hopfield82.pdf`.

Horn, D., Levy, N., & Ruppin, E. (1996). Neuronal–based synaptic compensation: a computational study in Alzheimer's disease. *Neural Computation 8*(6), 1227–1243.

Horn, D., Ruppin, E., Usher, M., & Hermann, M. (1993). Neural network modeling of memory deterioration in Alzheimer's disease. *Neural Computation 5*(5), 736–749.

Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural Networks 4*(2), 251–257.

Horvitz, E. J., Heckerman, D. E., & Langlotz, C. P. (1986). A framework for comparing alternative formalisms for plausible reasoning. *Proc. AAAI 5*, 210–214.

Hoxha, E., Boda, E., Montarolo, F., Parolisi, R., & Tempia, F. (2012). Excitability and synaptic alterations in the cerebellum of APP/PS1 mice. *PloS One 7*(4), e347265.

Hsiao, Y.-H., Chang, C.-H., & Gean, P.-W. (2018). Impact of social relationships on Alzheimer's memory impairment: mechanistic studies. *J Biomed Sci 25*(1), 3.

Hu, M.-L., Zong, X.-F., Mann, J. J., Zheng, J.-J., Liao, Y.-H., Li, Z.-C., He, Y., & Chen, X.-G. (2017). A review of the functional and anatomical default mode network in Schizophrenia. *Neuroscience Bulletin 33*(1), 73–84.

Huang, C.-C., Chung, C.-M., Leu, H.-B., Lin, L.-Y., Chiu, C.-C., Hsu, C.-Y., Chiang, C.-H., Huang, P.-H., Chen, T.-J., Lin, S.-J., Chen, J.-W., & Chan, W.-L. (2014). Diabetes mellitus and the risk of Alzheimer's disease: A nationwide population based study. *PloS One 9*(1), :e87095.

Humpel, C. (2011). Identifying and validating biomarkers for Alzheimer's disease. *Trends Biotechnol 29*(1), 26–32.

Hunter, S., & Brayne, C. (2017). Understanding the roles of mutations in the amyloid precursor protein in Alzheimer disease. *Molecular Psychiatry 23*, 23–81.

Huszár, F. (2017). Everything that works works because it's Bayesian: Why deep nets generalize? `http://www.inference.vc/everything-that-works-works-because-its-bayesian-2/`.

Ibata, K., Sun, Q., & Turrigiano, G. G. (2008). Rapid synaptic scaling induced by changes in postsynaptic firing. *Neuron 57*(6), 819–826.

Ikegaya, Y., Sasaki, T., Ishikawa, D., Honma, N., Tao, K., Takahashi, N., Minamisawa, G., Ujita, S., & Matsuki, N. (2013). Interpyramid spike transmission stabilizes the sparseness of recurrent network activity. *Cereb. Cortex 23*(2), 293–304.

Jaeger, H. (2002). Adaptive nonlinear system identification with echo state networks. *Proceedings of the 15th International Conference on Neural Information Processing Systems*, 609–616.

Jaeger, H., & Haas, H. (2004). Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication. *Science 304*(5667), 78–80.

Jang, S.-S., & Chung, H. J. (2016). Emerging link between Alzheimer's disease and homeostatic synaptic plasticity. *Neural Plasticity 1026*, 19 pages.

Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge: Cambridge University Press. Ed. by G. Larry Bretthorst. First publ. 1994. `https://archive.org/details/XQUHIUXHIQUHIQXUIHX2`, `http://www-biba.inrialpes.fr/Jaynes/prob.html`, `http://omega.albany.edu:8008/JaynesBook.html`.

Jeffrey, R. C. (Ed.) (1980). *Studies in inductive logic and probability. Vol. II*. Berkeley: University of California Press.

Jeffreys, H. (2003). *Theory of Probability* (3 ed.). London: Oxford University Press. First publ. 1939.

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage 17*(2), 825–841.

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., & Smith, S. M. (2012). Fsl. *NeuroImage 62*(2), 782–790.

Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis 5*(2), 143–156.

Jeong, H., Mason, S. P., Barabási, A.-L., & Oltvai, Z. N. (2001). Lethality and centrality in protein networks. *Nature 411*, 41–42.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., & Barabási, A.-L. (2000). The large-scale organization of metabolic networks. *Nature 407*, 651–654.

Johnson, G. V. W., & Stoothoff, W. H. (2004). Tau phosphorylation in neuronal cell function and dysfunction. *Journal of Cell Science 117*(24), 5721–5729.

Johnson, K. A., Fox, N. C., Sperling, R. A., & Klunk, W. E. (2012). Brain imaging in Alzheimer disease. *Cold Spring Harbor Perspectives in Medicine 2*(4), :a006213.

Johnson, N. L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika 36*(1/2), 149–176.

Jun, J. K., Miller, P., Hernández, A., Zainos, A., Lemus, L., Brody, C. D., & Romo, R. (2010). Heterogenous population coding of a short-term memory and decision task. *The Journal of Neuroscience 30*(3), 916–929.

Kadmon, J., & Sompolinsky, H. (2015). Transition to chaos in random neuronal networks. *Phys. Rev. X 5*(4), 041030.

Kallenberg, O. (2005). *Probabilistic Symmetries and Invariance Principles*. Probability and its applications. New York: Springer.

Kalna, G., & Higham, D. J. (2007). A clustering coefficient for weighted networks, with application to gene expression data. *AI Communications 20*(4), 263–271.

Kamenetz, F., Tomita, T., Hsieh, H., Seabrook, G., Borchelt, D., Iwatsubo, T., Sisodia, S., & Malinow, R. (2003). APP processing and synaptic function. *Neuron 37*(6), 925–937.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *J. Am. Stat. Assoc. 90*(430), 773–795. `https://www.stat.washington.edu/raftery/Research/PDF/kass1995.pdf`; `https://www.andrew.cmu.edu/user/kk3n/simplicity/KassRaftery1995.pdf`.

Keck, T., Keller, G. B., Jacobsen, R. I., Eysel, U. T., Bonhoeffer, T., & Hübener, M. (2013). Synaptic scaling and homeostatic plasticity in the mouse visual cortex invivo. *Neuron 80*(2), 327–334.

Keihaninejad, S., Heckemann, R. A., Fagiolo, G., Symms, M. R., Hajnal, J. V., & Hammers, A. (2010). A robust method to estimate the intracranial volume across MRI field strengths (1.5T and 3T). *NeuroImage 50*(4), 1427–37.

Kelleher, R. J., & Shen, J. (2015). Presenilin-1 mutations and Alzheimer's disease. *Proceedings of the National Academy of Sciences of the United States of America 114.4*, 629–631.

Khazaee, A., Ebrahimzadeh, A., & Babajani-Feremi, A. (2015). Identifying patients with Alzheimer's disease using resting-state fMRI and graph theory. *Clinical Neurophysiology 126*(11), 2132–41.

Khazaee, A., Ebrahimzadeh, A., & Babajani-Feremi, A. (2017). Classification of patients with MCI and AD from healthy controls using directed graph measures of resting-state fMRI. *Behavioural Brain Research 322*(Part B), 339–350.

Kim, H. K., Yoo, K., Na, D. L. N., Seo, S. W. S., Jeong, J., & Jeong, Y. (2015). Non-monotonic reorganization of brain networks with Alzheimer's disease progression. *Frontiers in Aging Neuroscience 7*, 111.

Kim, J., & Tsien, R. W. (2008). Synapse-specific adaptations to inactivity in hippocampal circuits achieve homeostatic gain control while dampening network reverberation. *Neuron 58*(6), 925–937.

Kingman, J. F. C. (1978). Uses of exchangeability. *Ann. Prob. 6*(2), 183–197.

Knight, R., Khondoker, M., Magill, N., Stewart, R., & Landau, S. (2018). A systematic review and meta-analysis of the effectiveness of acetylcholinesterase inhibitors and Memantine in treating the cognitive symptoms of dementia. *Dement Geriatr Cogn Disord 45*(3–4), 131–151.

Koch, G., & Spizzichino, F. (Eds.) (1982). *Exchangeability in Probability and Statistics*. Amsterdam: North-Holland.

Kolmogorov, A. N. (1942). Definition of center of dispersion and measure of accuracy to form a finite number of observations [Sur l'éstimation statistique des paramètres de la loi de Gauss]. *Izv. Akad. Nauk SSSR Ser. Mat. 6*(1–2), 3–32. In Russian.

Kook, J. H., Guindani, M., Zhang, L., & Vannucci, M. (2017). *NPBayes-fMRI*: Non-parametric Bayesian general linear models for single- and multi-subject fMRI data. *Stat. Biosci. 2017*, 1–19.

Koopman, B. O. (1936). On distributions admitting a sufficient statistic. *Trans. Am. Math. Soc. 39*(3), 399–409.

Koopman, B. O. (1940a). The axioms and algebra of intuitive probability. *Ann. Math. 41*(2), 269–292.

Koopman, B. O. (1940b). The bases of probability. *Bull. Am. Math. Soc. 46*, 763–774. Repr. in Kyburg & Smokler (1980), pp. 159–172.

Koopman, B. O. (1941). Intuitive probabilities and sequences. *Ann. Math. 42*(1), 169–187.

Koray, Ç. (2011). Minimum spanning tree reflects the alterations of the default mode network during Alzheimer's disease. *Annals of Biomedical Engineering 39*(5), 1493–1504.

Kotas, M. E., & Medzhitov, R. (2015). Homeostasis, inflammation, and disease susceptibility. *Cell 160*(5), 816–27.

Kotz, S., Balakrishnan, N., Read, C., Vidakovic, B., & Johnson, N. L. (Eds.) (2006). *Encyclopedia of Statistical Sciences* (2 ed.). Hoboken, USA: Wiley. First publ. 1982.

Kotz, S., & Nadarajah, S. (2004). *Multivariate t Distributions and Their Applications*. Cambridge: Cambridge University Press.

Kriener, B., Enger, H., Tetzlaff, T., Plesser, H. E., Gewaltig, M.-O., & Einevoll, G. T. (2014). Dynamics of self-sustained asynchronous-irregular activity in random networks of spiking neurons with strong synapses. *Front. Comput. Neurosci. 8*, 136.

Kruschwitz, J., List, D., Waller, L., Rubinov, M., & Walter, H. (2015). Graphvar: a user-friendly toolbox for comprehensive graph analyses of functional brain connectivity. *Journal of Neuroscience Methods 245*, 107–115.

Kurtz, D. S., & Swartz, C. W. (2004). *Theories of Integration: The Integrals of Riemann, Lebesgue, Henstock-Kurzweil, and Mcshane*, Volume 9 of *Series in real analysis*. Singapore: World Scientific.

Kyburg, Jr., H. E., & Smokler, H. E. (Eds.) (1980). *Studies in Subjective Probability* (2 ed.). Huntington, USA: Robert E. Krieger. First publ. 1964.

Lacor, P., Buniel, M., Furlow, P., Clemente, A., Velasco, P., Wood, M., Viola, K., & Klein, W. (2007). A$\beta$ oligomer-induced aberrations in synapse composition, shape, and density provide a molecular basis for loss of connectivity in Alzheimer's disease. *J. Neurosci. 27*(4), 796–807.

Lajoie, I., Nugent, S., Debacker, C., Dyson, K., Tancredi, F. B., Badhwar, A., Belleville, S., Deschaintre, Y., Bellec, P., Doyon, J., Bocti, C., Gauthier, S., Arnold, D., Kergoat, M.-J., Chertkow, H., Monchi, O., & Hoge, R. D. (2017). Application of calibrated fMRI in Alzheimer's disease. *NeuroImage: Clinical 15*, 348–358.

Lam, A. D., Deck, G., Goldman, A., Eskandar, E. N., Noebels, J., & Cole, A. J. (2017). Silent hippocampal seizures and spikes identified by foramen ovale electrodes in Alzheimer's disease. *Nature Medicine 23*, 678–680.

Lamoreaux, J., & Armstrong, G. (1998). The fundamental theorem of calculus for gauge integrals. *Math. Mag. 71*(3), 208–212.

Langton, C. G. (1990). Computation at the edge of chaos: phase transitions and emergent computation. *Physica D: Nonlinear Phenomena 42*(1-3), 12–37.

Lauritzen, S. L. (1988). *Extremal Families and Systems of Sufficient Statistics*, Volume 49 of *Lecture notes in statistics*. Berlin: Springer. First publ. 1982.

Lefort, S., Tomm, C., Floyd Sarria, J.-C., & Petersen, C. C. (2009). The excitatory neuronal network of the C2 barrel column in mouse primary somatosensory cortex. *Neuron 61*(2), 301–316.

Legenstein, R., & Maass, W. (2007a). Edge of chaos and prediction of computational performance for neural circuit models. *Neural Networks 20*(3), 323–334.

Legenstein, R., & Maass, W. (2007b). *What makes a dynamical system computationally powerful?*, pp. 127–154. MIT Press.

Leung, K. K., Barnes, J., Modat, M., Ridgway, G. R., Bartlett, J. W., Fox, N. C., & Ourselin, S. (2011). Brain maps: an automated, accurate and robust brain extraction technique using a template library. *NeuroImage 55*(3), 1091–108.

Levine, R. D., & Tribus, M. (Eds.) (1979). *The Maximum Entropy Formalism: A Conference Held at the Massachusetts Institute of Technology on May 2–4, 1978*. Cambridge, USA: MIT Press.

Li, G., Ramanathan, K., Ning, N., Shi, L., & Wen, C. (2015). Memory dynamics in attractor networks. *Comput Intell Neurosci 2015*, 191745.

Lin, J.-W., & Faber, D. S. (2002). Modulation of synaptic delay during synaptic plasticity. *Trends Neurosci. 25*(9), 449–455.

Lindley, D. V. (2008). *Introduction to Probability and Statistics from a Bayesian Viewpoint. Part 2: Inference* (reprint ed.). Cambridge: Cambridge University Press. First publ. 1965.

Lindley, D. V., & Phillips, L. D. (1976). Inference for a Bernoulli process (a Bayesian view). *American Statistician 30*(3), 112–119.

Lindquist, M. A. (2008). The statistical analysis of fMRI data. *Stat. Sci. 23*(4), 439–464.

Lissin, D. V., Gomperts, S. N., Carroll, R. C., Christine, C. W., Kalman, D., Kitamura, M., Hardy, S., Nicoll, R. A., Malenka, R. C., & von Zastrow, M. (1998). Activity differentially regulates the surface expression of synaptic AMPA and NMDA glutamate receptors. *Proc Natl Acad Sci U S A 95*(12), 7097–7102.

Little, W. (1974). The existence of persistent states in the brain. *Mathematical Biosciences 10*(1), 101–120.

Liu, Q., Xie, X., Lukas, R. J., St. John, P. A., & Wu, J. (2013). A novel nicotinic mechanism underlies $\beta$-amyloid-induced neuronal hyperexcitation. *J Neurosci 33*(17), 7253–63.

Liu, X., Gerraty, R. T., Grinband, J., Parker, D., & Razlighi, Q. R. (2017). Brain atrophy can introduce age-related differences in bold response. *Human Brain Mapping 38*(7), 3402–3414.

Lloyd, D., Aon, A. M., & Cortassa, S. (2001). Why homeodynamics, not homeostasis? *TheScientificWorldJOURNAL 1*, 133–145.

London, M., Roth, A., Beeren, L., Häusser, M., & Latham, P. E. (2010). Sensitivity to perturbations in vivo implies high noise and suggests rate coding in cortex. *Nature 466*(1), 123–128.

Lu, Y., Jiang, T., & Zang, Y.-F. (2003). Region growing method for the analysis of functional MRI data. *NeuroImage 20*(1), 455–65.

Lütcke, H., Margolis, D. J., & Helmchen, F. (2013). Steady or changing? Long-term monitoring of neuronal population activity. *Trends in Neurosciences 36*(7), 375–384.

Ma, H., & Zeng, A.-P. (2003). The connectivity structure, giant strong component and centrality of metabolic networks. *Bioinformatics 19 11*, 1423–30.

Maass, W., Natschläger, T., & Markram, H. (2002). Real-time computing without stable states: a new framwork for neural compuation based on perturbation. *Neural Comput. 14*(11), 2531–2560.

MacKay, D. J. C. (1992a). Bayesian interpolation. *Neural Comp. 4*(3), 415–447. `http://www.inference.phy.cam.ac.uk/mackay/PhD.html`.

MacKay, D. J. C. (1992b). *Bayesian Methods for Adaptive Models*. Ph. D. thesis, California Institute of Technology, Pasadena, USA. http://www.inference.phy.cam.ac.uk/mackay/PhD.html.

MacKay, D. J. C. (1992c). The evidence framework applied to classification networks. *Neural Comp. 4*(5), 720–736. http://www.inference.phy.cam.ac.uk/mackay/PhD.html.

MacKay, D. J. C. (1992d). Information-based objective functions for active data selection. *Neural Comp. 4*(4), 590–604. http://www.inference.phy.cam.ac.uk/mackay/PhD.html.

MacKay, D. J. C. (1992e). A practical Bayesian framework for backpropagation networks. *Neural Comp. 4*(3), 448–472. http://www.inference.phy.cam.ac.uk/mackay/PhD.html.

MacKay, D. J. C. (2003). *Information Theory, Inference, and Learning Algorithms*. Cambridge: Cambridge University Press. http://www.inference.phy.cam.ac.uk/mackay/itila/. First publ. 1995.

Maclure, M. (1998). How to change your mind. *Epidemiology 9*(3), 233. See Greenland (1998).

Major, G., Larkum, M., & Schiller, J. (2013). Active properties of neocortical pyramidal neuron dendrites. *Annu. Rev. Neurosci. 36*, 1–24.

Marcus, C., Mena, E., & Subramaniam, R. M. (2014). Brain pet in the diagnosis of Alzheimer's disease. *Clin Nucl Med 39*(10), e413–e426.

Marder, E., & Goaillard, J.-M. (2006). Variability, compensation and homeostasis in neuron and network function. *Nature Reviews Neuroscience 7*, 563—-574.

Marrelec, G., & Fransson, P. (2011). Assessing the influence of different ROI selection strategies on functional connectivity analyses of fMRI data acquired during steady-state conditions. *PLoS One 6*(4), e14788.

Masliah, E., Mallory, M., Hansen, L., DeTeresa, R., Alford, M., & Terry, R. (1994). Silent hippocampal seizures and spikes identified by foramen ovale electrodes in Alzheimer's disease. *Neuroscience Letters 174*, 67–72.

Mayeux, R., & Stern, Y. (2012). Epidemiology of Alzheimer disease. *Cold Spring Harb Perspect Med 2*(8), :a006239.

Mboup, M., & Larsen, T. A. É. M. J. (Eds.) (2014). *2014 IEEE International Workshop on Machine Learning for Signal Processing*. New York: IEEE.

McCarthy, J. (1956). Measures of the value of information. *Proc. Natl. Acad. Sci. (USA) 42*(9), 654–655.

McCullagh, P. (2002). What is a statistical model? *Ann. Stat. 30*(5), 1225–1267. http://www.stat.uchicago.edu/~pmcc/publications.html. See also the following discussion and rejoinder Besag et al. (2002).

McGuire, B. A., Gilbert, C. D., Rivlin, P. K., & Wiesel, T. N. (1991). Targets of horizontal connections in macaque primary visual cortex. *Journal of Comparative Neurology 305*(3), 370–392.

McKenzie, K. J. (2014). How does untreated psychosis lead to neurological damage? *Can. J. Psychiatry 59*(10), 511–512.

Mead, R. (1965). A generalised logit-normal distribution. *Biometrics 21*(3), 721–732.

Mendez, M. F., Catanzaro, P., Doss, R. C., Arguello, R., & Frey, W. H. (1994). Seizures in Alzheimer's disease: clinicopathologic study. *Journal of Geriatric Psychiatry and Neurology 7*(4), 0891–9887.

Menkes-Caspi, N., Yamin, H. G., Kellner, V., Spires-Jones, T. L., Cohen, D., & Stern, E. A. (2015). Pathological tau disrupts ongoing network activity. *Neuron 85*(5), 959–966.

Merino-Serrais, P., Benavides-Piccione, R., Blazquez-Llorca, L., Kastanauskaite, A., Rábano, A., Avila, J., & DeFelipe, J. (2013). The influence of phospho-tau on dendritic spines of cortical pyramidal neurons in patients with Alzheimer's disease. *Brain 136*(6), 1913–1928.

Michael, A. M., Baum, S. A., White, T., Demirci, O., Andreasen, N. C., Segall, J. M., Jung, R. E., Pearlson, G., Clark, V. P., Gollub, R. L., Schulz, S. C., Roffman, J. L., Lim, K. O., Ho, B.-C., Bockholt, H. J., & Calhoun, V. D. (2010). Does function follow form?: Methods to fuse structural and functional brain images show decreased linkage in Schizophrenia. *NeuroImage 49*(3), 2626–2637.

Minka, T. (2001). Inferring a Gaussian distribution. Technical report, MIT media Lab, Cambridge, USA. `http://research.microsoft.com/en-us/um/people/minka/papers/`. First publ. 1998.

Morris, J. C. (2005). Early-stage and preclinical Alzheimer disease. *Alzheimer Dis Assoc Disord 19*, 163–165.

Morrison, A., Straube, S., Plesser, H. E., & Diesmann, M. (2007). Exact subthreshold integration with continuous spike times in discrete-time neural network simulations. *Neural Computation 19*(1), 47–49.

Morton, A. J., Faull, R. L. M., & Edwardson, J. M. (2001). Abnormalities in the synaptic vesicle fusion machinery in Huntington's diseasee. *Brain Research Bulletin 56*(2), 111–117.

Mosteller, F., & Wallace, D. L. (1963). Inference in an authorship problem: A comparative study of discrimination methods applied to the authorship of the disputed *federalist* papers. *J. Am. Stat. Assoc. 58*(302), 275–309. `https://www.stat.cmu.edu/Exams/mosteller.pdf`.

Murphy, K., Bodurka, J., & Bandettini, P. A. (2007). How long to scan the relationship between fMRI temporal signal to noise and necessary scan duration. *NeuroImage 34*(2), 565–574.

Murphy, K., & Fox, M. D. (2017). Towards a consensus regarding global signal regression for resting state functional connectivity MRI. *NeuroImage 154*, 169–173.

Murphy, K. P. (2007). Conjugate Bayesian analysis of the Gaussian distribution. `http://thaines.com/content/misc/gaussian_conjugate_prior_cheat_sheet.pdf`.

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. Adaptive computation and machine learning series. Cambridge, USA: MIT Press.

Murray, J. D., Bernacchia, A., Roy, N. A., Constantinidis, C., Romo, R., & Wang, X.-J. (2017). Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. *Proc. Natl. Acad. Sci. USA 114*(2), 394–399.

Naudé, J., Cessac, B., Berry, H., & Delord, B. (2013). Effects of cellular homeostatic intrinsic plasticity on dynamical and computational properties of biological recurrent neural networks. *J. Neurosci. 33*(38), 15032–15043.

Nelson, P. T., Alafuzoff, I., Bigio, E. H., Bouras, C., Braak, H., Cairns, N. J., Castellani, R. J., Crain, B. J., Davies, P., Tredici, K. D., Duyckaerts, C., Frosch, M. P., Haroutunian, V., Hof, P. R., Hulette, C. M., Hyman, B. T., Iwatsubo, T., Jellinger, K. A., Jicha, G. A., Kövari, E., Kukull, W. A., Leverenz, J. B., Love, S., Mackenzie, I. R., Mann, D. M., Masliah, E., McKee, A. C., Montine, T. J., Morris, J. C., Schneider, J. A., Sonnen, J. A., Thal, D. R., Trojanowski, J. Q., Troncoso, J. C., Wisniewski, T., Woltjer, R. L., & Beach, T. G. (2012). Correlation of Alzheimer disease neuropathologic changes with cognitive status: a review of the literature. *Journal of Neuropathology & Experimental Neurology 71*(5), 362–381.

Neuman, K. M., Molina-Campos, E., Musial, T. F., Price, A. L., Oh, K.-J., Wolke, M. L., Buss, E. W., Scheff, S. W., Mufson, E. J., & Nicholson, D. A. (2014). Evidence for Alzheimer's disease-linked synapse loss and compensation in mouse and human hippocampal CA1 pyramidal neurons. *Brain Struct Funct 220*(6), 3143–65.

Newman, M. E. J. (2004). Analysis of weighted networks. *Physical review. E, Statistical, nonlinear, and soft matter physics 70*(5 PT 2).

Neyman, J. (1935). Su un teorema concernente le cosiddette statistiche sufficienti. *Giorn. Ist. Ital. Att. VI*(4), 320–334.

Nielsen, S. F. V., Madsen, K. H., Røge, R., Schmidt, M. N., & Mørup, M. (2016). Nonparametric modeling of dynamic functional connectivity in fMRI data. `arXiv:1601.00496`.

Nordlie, E., Gewaltig, M.-O., & Plesser, H. E. (2009). Towards reproducible descriptions of neuronal network models. *PLOS Comput. Biol. 5*(8), e1000456.

Nordlie, E., Tetzlaff, T., & Einevoll, G. T. (2010). Rate dynamics of leaky integrate-and-fire neurons with strong synapses. *Front. Comput. Neurosci. 4*, 149.

O'Brien, J., O'Keefe, K., LaViolette, P., DeLuca, A., Blacker, D., Dickerson, B., & Sperling, R. (2010). Longitudinal fMRI in elderly reveals loss of hippocampal activation with clinical decline. *Neurology 75*(24), 1969–1976.

O'Brien, R. J., Kamboj, S., Ehlers, M. D., Rosen, K. R., Fischbach, G. D., & Huganir, R. L. (1998). Activity-dependent modulation of synaptic AMPA receptor accumulation. *Neuron 21*, 1067–1078.

Okun, M., & Lampl, I. (2008). Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. *Nat. Neurosci. 11*(5), 535–537.

Orbán, G., Völgyi, K., Juhász, G., Penke, B., Kékesi, K. A., Kardos, J., & Czurkó, A. (2010). Different electrophysiological actions of 24- and 72-hour aggregated amyloid-beta oligomers on hippocampal field population spike in both anesthetized and awake rats. *Brain Research 1354*(6), 227–235.

Ostojic, S. (2014). Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons. *Nat. Neurosci. 17*, 594–600.

Ostojic, S. (2015). Response to comment on "Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons". *bioRxiv*, 020354.

Pagani, M., Giuliani, A., Öberg, J., Chincarini, A., Morbelli, S., Brugnolo, A., Arnaldi, D., Picco, A., Bauckneht, M., Buschiazzo, A., Sambuceti, G., & Nobili, F. (2016). Predicting the transition from normal aging to Alzheimer's disease: a statistical mechanistic evaluation of FDG-PET data. *NeuroImage 141*, 282–290.

Pagani, M., Giuliani, A., Öberg, J., De Carli, F., Morbelli, S., Girtler, N., Arnaldi, D., Accardo, J., Bauckneht, M., Bongioanni, F., Chincarini, A., Sambuceti, G., Jonsson, C., & Nobili, F. (2017). Progressive disintegration of brain networking from normal aging to Alzheimer disease. *Journal of Nuclear Medicine 58*, 1132–1139.

Paris, J. B. (2006). *The Uncertain Reasoner's Companion: A Mathematical Perspective* (reprint ed.). Number 39 in Cambridge tracts in theoretical computer science. Cambridge: Cambridge University Press. Concerning Cox's theorem see also Snow (1998).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: machine learning in Python. *Journal of Machine Learning Research 12*, 2825–2830.

Pereda, E., Quian Quiroga, R., & Bhattacharya, J. (2005). Nonlinear multivariate analysis of neurophysiological signals. *Progress in Neurobiology 77*(1–2), 1–37.

Pereira, U., & Brunel, N. (2018). Attractor dynamics in networks with learning rules inferred from in vivo data. *Neuron 99*(1), 277–238.e4.

Perez, C., Ziburkus, J., & Ullah, G. (2016). Analyzing and modeling the dysfunction of inhibitory neurons in Alzheimer's disease. *PloS One 11*(12), :e0168800.

Petersen, C. C., & Crochet, S. (2013). Synaptic computation and sensory processing in neocortical layer 2/3. *Neuron 78*(1), 28–48.

Pini, L., Pievani, M., Bocchetta, M., Altomare, D., Bosco, P., Cavedo, E., Galluzzi, S., Marizzoni, M., & Frisoni, G. B. (2016). Brain atrophy in Alzheimer's disease and aging. *Ageing Research Reviews 30*, 25–48.

Pitman, E. J. G. (1936). Sufficient statistics and intrinsic accuracy. *Math. Proc. Camb. Phil. Soc. 32*(4), 567–579.

Poisnel, G., Herard, A.-S., El Tannir El Tayara, N., Bourrin, E., Volk, A., Kober, F., Delatour, B., Delzescaux, T., Debeir, T., Rooney, T., Benavides, J., Hantraye, P., & Dhenain, M. (2012). Increased regional cerebral glucose uptake in an APP/PS1 model of Alzheimer's disease. *Neurobiology of Aging 33*(9), 1995–2005.

Pólya, G. (1949). Preliminary remarks on a logic of plausible inference. *Dialectica 3*(1–2), 28–35.

Pólya, G. (1968). *Mathematics and Plausible Reasoning: Vol. II: Patterns of Plausible Inference* (2 ed.). Princeton: Princeton University Press. First publ. 1954.

Pope, S. K., Shue, V. M., & Beck, C. (2003). Will a healthy lifestyle help prevent Alzheimer's disease? *Annual Reviews 24*(1), 111–132.

Popescu, V., Battaglini, M., Hoogstrate, W., Verfaillie, S., Sluimer, I., van Schijndel, R., van Dijk, B., Cover, K., Knol, D., Jenkinson, M., Barkhof, F., de Stefano, N., & Vrenken, H. (2012). Optimizing parameter choice for FSL-brain extraction tool (BET) on 3D T1 images in multiple sclerosis. *NeuroImage 61*(4), 1484–1494.

Porta Mana, P. G. L. (2017). Model comparison and Bayes factors: what is a model? In preparation.

Porta Mana, P. G. L., Bachmann, C., & Morrison, A. (2018). Inferring health conditions from fMRI-graph data. Open Science Framework doi:10.17605/osf.io/r2huz.

Raiffa, H., & Schlaifer, R. (2000). *Applied Statistical Decision Theory* (reprint ed.). Wiley Classics Library. New York: Wiley. First publ. 1961.

Rajan, K., & Abbott, L. F. (2006). Eigenvalue spectra of random matrices for neural networks. *Phys. Rev. Lett. 97*, 188104.

Rajendran, L., & Paolicelli, R. C. (2018). Microglia–mediated synapse loss in Alzheimer's disease. *The Journal of Neuroscience 38*(12), 2911–2919.

Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., & Barabási, A.-L. (2002). Hierarchical organization of modularity in metabolic networks. *Science 297*(5586), 1551–1555.

Reese, J. P., Heßmann, P., Seeberg, G., Henkel, D., Hirzmann, P., Rieke, J., Baum, E., Dannhoff, F., Müller, M. J., Jessen, F., Geldsetzer, M.-B., & Dodel, R. (2011). Cost and care of patients with Alzheimer's disease: clinical predictors in German health care settings. *Journal of Alzheimer's Disease 27*(4), 723–736.

Reger, M. A., Henderson, S. T., Hale, C., Cholerton, B., Baker, L. D., Watson, G., Hyde, K., Chapman, D., & Craft, S. (2004). Effects of $\beta$-hydroxybutyrate on cognition in memory-impaired adults. *Neurobiology of Aging 25*(3), 311–314.

Rivest, R. L., Leiserson, C. E., & Cormen, T. H. (2000). *Introduction to Algorithms (Mit Electrical Engineering and Computer Science Series.)*. Cambridge, Mass.: MIT Press.

Rizzuto, R., De Stefani, D., Raffaello, & Anna, Mammucari, C. (2012). Mitochondria as sensors and regulators of calcium signalling. *Nature Reviews Molecular Cell Biology 13*, 566–578.

Rodgers, J. L., & Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician 42*(1), 59–66.

Rotter, S., & Diesmann, M. (1999). Exact digital simulation of time-invariant linear systems with applications to neuronal modeling. *Biol. Cybern. 81*(5-6), 381–402.

Ruppin, E., & Reggia, J. A. (1994). A neural model of memory impairment in diffuse cerebral atrophy. *The British Journal of Psychiatry 166*(1), 19–28.

Santhakumar, V., & Soltesz, I. (2004). Plasticity of interneuronal species diversity and parameter variance in neurological diseases. *Trends in Neurosciences 27*(8), 504–510.

Sanz-Arigita, E. J., Schoonheim, M. M., Damoiseaux, J. S., Rombouts, S. A., Maris, E., Barkhof, F., Scheltens, P., & Stam, C. J. (2010). Loss of "small-world" networks in Alzheimer's disease: graph analysis of fMRI resting-state functional connectivity. *PloS One 5*(11), :e13788.

Schaeffer, S. E. (2007). Graph clustering. *Computer Science Review 1*(1), 27 – 64.

Scheff, S. W., & Price, D. A. (2003). Synaptic pathology in Alzheimers disease: a review of ultrastructural studies. *Neurobiology of Aging 24*(8), 1029 – 1046. Molecular and Cellular Basis of Synaptic Loss and Dysfunction in Alzheimer's Disease.

Scheff, W. S., DeKosky, S. T., & Price, D. A. (1990). Quantitative assessment of cortical synaptic density in Alzheimer's disease. *Neurobiology of Aging 11*(1), 29–37.

Scheff, W. S., & Douglas, A. P. (2006). Alzheimer's disease-related alterations in synaptic density: neocortex and hippocampus. *Journal of Alzheimer's disease: JAD 9*, 101–115.

Scheff, W. S., Douglas, A. P., Schmitt, F. A., Scheff, M. A., & Mufson, E. J. (2011). Synaptic loss in the inferior temporal gyrus in mild cognitive impairment and Alzheimer disease. *J Alzheimers Dis 24*(3), 547–557.

Scheff, W. S., Neltner, J. H., & Nelson, P. T. (2014). Is synaptic loss a unique hallmark of Alzheimer's disease? *Biochem Pharmacol 88*(4), 517–528.

Scheff, W. S., & Price, D. A. (1993). Synapse loss in the temporal lobe in Alzheimer's disease. *Ann Neurol 33*(2), 190–199.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks 61*, 85–117.

Scholl, D. A. (1956). *THE ORGANIZATION OF THE CEREBRAL CORTEX*. New York: Springer-Verlag.

Schrauwen, B., Büsing, L., & Legenstein, R. A. (2009). On computational power and the order-chaos phase transition in reservoir computing. In *Advances in Neural Information Processing Systems*, pp. 1425–1432.

Schreiber, T. (2000). Measuring information transfer. *Phys Rev Lett. 85*(2), 461–4.

Schroeter, M. L., Timo, S., Maslowski, N., & Neumann, J. (2009). Neural correlates of Alzheimer's disease and mild cognitive impairment: a systematic and quantitative meta-analysis involving 1,351 patients. *NeuroImage 47*(4), 1196–1206.

Schuecker, J., Diesmann, M., & Helias, M. (2015). Modulated escape from a metastable state driven by colored noise. *Phys. Rev. E 92*(5), 052119.

Schuecker, J., Goedeke, S., & Helias, M. (2017). Optimal sequence memory in driven random networks. *arXiv*. 1603.01880v3 [q-bio.NC].

Schwarz, A. J., & McGonigle, J. (2011). Negative edges and soft thresholding in complex network analysis of resting state functional connectivity data. *NeuroImage 55*(3), 1032–1146.

Shadlen, M. N., & Newsome, W. T. (1998). The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J. Neurosci. 18*(10), 3870–3896.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J. 27*(3, 4), 379–423, 623–656. https://archive.org/details/bstj27-3-379, https://archive.org/details/bstj27-4-623, http://math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf.

Shao, J., Myers, N., Yang, Q., Feng, J., Plant, C., Böhm, C., Förstl, H., Kurz, A., Meng, C., Riedl, V., Wohlschläger, A., & Sorg, C. (2012). Prediction of Alzheimer's disease using individual structural connectivity networks. *Neurobiol Aging 33*(12), 2756–2765.

Shen, J., & Kelleher, R. J. (2007). The presenilin hypothesis of Alzheimer's disease: evidence for a loss-of-function pathogenic mechanism. *Proceedings of the National Academy of Sciences of the United States of America 104*(2), 403–409.

Sheng, M., Sabatini, B., & Südhof, T. C. (2012). Synapses and Alzheimer's disease. *Cold Spring Harb Perspect Biol 4*(5), pii:a005777.

Shenton, M. E., Whitford, T. J., & Kubicki, M. (2010). Structural neuroimaging in Schizophrenia from methods to insights to treatments. *Dialogues in Clinical Neuroscience 12*(3), 317–332.

Shergill, S., Brammer, M., Williams, S., Murray, R., & McGuire, P. (2000). Mapping auditory hallucinations in Schizophrenia using functional magnetic resonance imaging. *Archives of General Psychiatry 57*(11), 1033–1038.

Silva, R. F., Castro, E., Gupta, C. N., Cetin, M., Arbabshirani, M., Potluru, V. K., Plis, S. M., & Calhoun, V. D. (2014). The tenth annual MLSP competition: Schizophrenia classification challenge. In Mboup & Larsen (2014), pp. 6958889.

Slomowitz, E., B, S., Vertkin, I., Milshtein-Parush, H., Nelken, I., Slutsky, M., & Slutsky, I. (2015). Interplay between population firing stability and single neuron dynamics in hippocampal networks. *eLife 4*, e04378.

Small, D. H. (2004). Mechanisms of synaptic homeostasis in Alzheimer's disease. *Current Alzheimer Research 1*(1), 27–32.

Smith, A. (2002a). Effects of caffeine on human behavior. *Food Chem. Toxicol. 40*(9), 1243–1255. http://intraspec.ca/Effects-of-caffeine-on-human-health.pdf.

Smith, J. C., Nielson, K. A., Antuono, P., Lyons, J.-A., Hanson, R. J., Butts, A. M., Hantke, N. C., & Verber, M. D. (2013). Semantic memory fMRI and cognitive function after exercise intervention in mild cognitive impairment. *J Alzheimers Dis 37*(1), 197–215.

Smith, S. M. (2002b). Fast robust automated brain extraction. *Hum. Brain Mapp. 17*(3), 143–155.

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., Bannister, P. R., De Luca, M., Drobnjak, I., Flitney, D. E., Niazy, R. K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J. M., & Matthews, P. M. (2005). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage 23*(Suppl 1), S208–219.

Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., Ramsey, J. D., & Woolrich, M. W. (2011). Network modelling methods for fMRI. *NeuroImage 54*(2), 875–891.

Snow, P. (1998). On the correctness and reasonableness of Cox's theorem for finite domains. *Comput. Intell. 14*(3), 452–459.

Softky, W. R., & Koch, C. (1993). The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *J. Neurosci. 13*(1), 334–350.

Sompolinsky, H., Crisanti, A., & Sommers, H. J. (1988). Chaos in random neural networks. *Phys. Rev. Lett. 61*(3), 259–262.

Song, S., Sjöström, P., Reigl, M., Nelson, S., & Chklovskii, D. (2005). Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLOS Biol. 3*(3), e68.

Sox, H. C., Higgins, M. C., & Owens, D. K. (2013). *Medical Decision Making* (2 ed.). New York: Wiley. First publ. 1988.

Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., Iwatsubo, T., Jack, C. R. J., Kaye, J., Montine, T. J., Park, D. C., Reiman, E. M., Rowe, C. C., Siemers, E., Stern, Y., Yaffe, K., Carrillo, M. C., Thies, B., Morrison-Bogorad, M., Wagster, M. V., & Phelps, C. H. (2011). Toward defining the preclinical stages of Alzheimer's disease: recommendations from the national institute on aging–Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's and Dementia 7*(3), 280–292.

Stam, C. J., & Reijneveld, J. C. (2007). Graph theoretical analysis of complex networks in the brain. *Nonlinear Biomedical Physics 1*, :3.

Steriade, M., Nuñez, A., & Amzica, F. (1993). A novel slow ($< 1$ Hz) oscillation of neocortical neurons in vivo: depolarizing and hyperpolarizing components. *J. Neurosci. 13*(8), 3252–3265.

Stern, Y. (2012). Cognitive reserve in ageing and Alzheimer's disease. *Lancet Neurol 11*(11), 1006–12.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *J. Roy. Stat. Soc. B 36*(2), 111–133. See also discussion Barnard et al. (1974).

Styr, B., & Slutsky, I. (2018). Imbalance between firing homeostasis and synaptic plasticity drives early-phase Alzheimer's disease. *Nature Neuroscience 21*(4), 463–473.

Sulmasy, D. P., Morgan, T., Caubet, J.-F., & Goodman, S. (2000). Toward evidence-based statistics [comments and response]. *Ann. Intern. Med. 132*(6), 507–508. See Goodman (1999); Davidoff (1999).

Supekar, K., Menon, V., Rubin, D., Musen, M., & Greicius, M. D. (2008). Network analysis of intrinsic functional brain connectivity in Alzheimer's disease. *PLoS Computational Biology 4*(6), :e1000100.

Sussillo, D., & Abbott, L. F. (2009). Generating coherent patterns of activity from chaotic neural networks. *Neuron 63*(4), 544–557.

Swartz, C. (2001). *Introduction to Gauge Integrals*. Singapore: World Scientific.

Swensen, A. M., & Bean, B. P. (2005). Robustness of burst firing in dissociated Purkinje neurons with acute or long-term reductions in sodium conductance. *Journal of Neuroscience 25*(14), 3509–3520.

Takeuchi, T., Duszkiewicz, A. J., & Morris, R. G. M. (2014). The synaptic plasticity and memory hypothesis: encoding, storage and persistence. *Philos Trans R Soc Lond B Biol Sci. 369*(1633), 20130288.

Tampellini, D. (2015). Synaptic activity and Alzheimer's disease a critical update. *Frontiers in Neuroscience 9*, 432–439.

Telesford, Q., Morgan, A., Hayasaka, S., Simpson, S., Barret, W., Kraft, R., Mozolic, J., & Laurienti, P. (2010). Reproducibility of graph metrics in fMRI networks. *Frontiers in Neuroinformatics 4*, 117.

Terenin, A., & Draper, D. (2017). Cox's theorem and the Jaynesian interpretation of probability. `arXiv:1507.06597`. First publ. 2015.

Terry, R. D., Masliah, E., Salmon, D. P., Butters, N., DeTeresa, R., Hill, R., Hansen, L. A., & Katzman, R. (1991). Physical basis of cognitive alterations in Alzheimer's disease: synapse loss is the major correlate of cognitive impairment. *Annals of Neurology 30*(4), 572–580.

Tetzlaff, T., Helias, M., Einevoll, G. T., & Diesmann, M. (2012). Decorrelation of neural-network activity by inhibitory feedback. *PLOS Comput. Biol. 8*(8), e1002596.

ISO (2006). *ISO 3534-1:2006: Statistics – Vocabulary and symbols – Part 1: General statistical terms and terms used in probability*. Geneva: International Organization for Standardization.

ISO (2009). *ISO 80000:2009: Quantities and units*. Geneva: International Organization for Standardization. First publ. 1993.

Thalmeier, D., Uhlmann, M., Kappen, H. J., & Memmesheimer, R.-M. (2016). Learning universal computations with spikes. *PLOS Computational Biology 12*(6), :e1004895.

Thiagarajan, T. C., Lindskog, M., & Tsien, R. W. (2005). Adaptation to synaptic inactivity in hippocampal neurons. *Neuron 47*(5), 725–37.

Thirion, B., Varoquaux, G., Dohmatob, E., & Poline, J.-B. (2014). Which fMRI clustering gives good brain parcellations? *Frontiers in Neuroscience 8*(167), 167.

Thompson, P. M., Hayashi, Kiralee M.and de Zubicaray, G., Janke, A. L., Rose, S. E., Semple, J., Herman, D., Hong, M. S., Dittmer, S. S., Doddrell, D. M., & Toga, A. W. (2003). Dynamics of gray matter loss in Alzheimer's disease. *The Journal of Neuroscience 23*, 994–1005.

Tiao, G. C., & Zellner, A. (1964). On the Bayesian estimation of multivariate regression. *J. Roy. Stat. Soc. B 26*(2), 277–285.

Timofeev, I., Grenier, F., & Steriade, M. (2001). Disfacilitation and active inhibition in the neocortex during the natural sleep-wake cycle: an intracellular study. *Proc. Natl. Acad. Sci. USA 98*(4), 1924–1929.

Tomko, G. J., & Crapper, D. R. (1974). Neuronal variability: non-stationary responses to identical visual stimuli. *Brain Research 79*(3), 405–418.

Tönnies, E., & Trushina, E. (2017). Oxidative stress, synaptic dysfunction, and Alzheimer's disease. *Journal of Alzheimer's disease 57*(4), 1105–1121.

Trombetta, B. A., Carlyle, B. C., Koenig, A. M., Shaw, L. M., Trojanowski, J. Q., Wolk, D. A., Locascio, J. J., & Arnold, S. E. (2018). The technical reliability and biotemporal stability of cerebrospinal fluid biomarkers for profiling multiple pathophysiologies in Alzheimer's disease. *PloS One 13*(3), e0193707.

Turrigiano, G. G. (2008). The self-tuning neuron: synaptic scaling of excitatory synapses. *Cell 135*(3), 422–435.

Turrigiano, G. G. (2012). Homeostatic synaptic plasticity: local and global mechanisms for stabilizing neuronal function. *Cold Spring Harb Perspect Biol 4*(1), a005736.

Turrigiano, G. G., Leslie, K. R., Desai, N. S., Rutherford, L. C., & Nelson, S. B. (1998). Activity-dependent scaling of quantal amplitude in neocortical neurons. *Nature 391*, 892–896.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., & Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage 15*(1), 273–289.

van Albada, S. J., Helias, M., & Diesmann, M. (2015). Scalability of asynchronous networks is limited by one-to-one mapping between effective connectivity and correlations. *PLOS Comput. Biol. 11*(9), e1004490.

van den Oord, A., Kalchbrenner, N., & Kavukcuoglu, K. (2016). Generative adversarial networks. *arXiv arXiv:1406.2661 [stat.ML]*.

Van Welie, I., van Hooft, J. A., & Wadman, W. J. (2004). Homeostatic scaling of neuronal excitability by synaptic modulation of somatic hyperpolarization-activated Ih channels. *Proceedings of the National Academy of Sciences of the United States of America 101*(14), 5123–5128.

Vemuri, P., Jones, D. T., & Jack, C. R. (2012). Resting state functional MRI in Alzheimer's disease. *Alzheimer's Research & Therapy 4*(1), 2.

Venkataraman, A., Whitford, T. J., Westin, C.-F., Golland, P., & Kubicki, M. (2012). Whole brain resting state functional connectivity abnormalities in schizophrenia. *Schizophr. Res. 139*(1–3), 7–12.

Venturelli, M., Scarsini, R., & Schena, F. (2011). Six-month walking program changes cognitive and ADL performance in patients with Alzheimer. *Am J Alzheimers Dis Other Demen 26*(5), 381–388.

Vitureira, N., & Goda, Y. (2013). The interplay between Hebbian and homeostatic synaptic plasticity. *J Cell Biol 203*(2), 175–186.

Vitureira, N., Letellier, M., & Goda, Y. (2012). Homeostatic synaptic plasticity: from single synapses to neural circuits. *Current Opinion in Neurobiology 22*(3), 516–521.

Šiškova, Z., Justus, D., Kaneko, H., Friedrichs, D., Henneberg, N., Beutel, T., Pitsch, J., Schoch, S., Becker, A., von der Kammer, H., & Remy, S. (2014). Dendritic structural degeneration is functionally linked to cellular hyperexcitability in a mouse model of Alzheimer's disease. *Neuron 84*, 1023–1033.

Wagner, C., & Stucki, J. W. (2002). Construction of an associative memory using unstable periodic orbits of a chaotic attractor. *Journal of Theoretical Biology 215*(03), 375–384.

Wang, H. E., Bénar, C. G., Quilichini, P. P., Friston, K. J., Jirsa, V. K., & Bernard, C. (2014). A systematic framework for functional connectivity measures. *Frontiers in Neuroscience 8*, 405.

Wang, Z., Zhang, M., Han, Y., Song, H., Guo, R., & Li, K. (2016). Differentially disrupted functional connectivity of the subregions of the amygdala in Alzheimer's disease . *Journal of X-Ray Science and Technology 24*(2), 329–342.

Watt, A. J., van Rossum, M. C., MacLeod, K. M., Nelson, S. B., & Turrigiano, G. G. (2000). Activity coregulates quantal AMPA and NMDA currents at neocortical synapses. *Neuron 26*(3), 659–679.

Wattamwar, P. R., & Mathuranath, P. S. (2010). An overview of biomarkers in Alzheimer's disease. *Ann Indian Acad Neurol 13*(Suppl2), S116–23.

Weintraub, S., Wicklund, A. H., & Salmon, D. P. (2012). The neuropsychological profile of Alzheimer disease. *Cold Spring Harb Perspect Med 2*(4), a006171.

Wieland, S., Bernardi, D., Schwalger, T., & Lindner, B. (2015). Slow fluctuations in recurrent networks of spiking neurons. *Physical Review E 92*(4), 040901.

Woodward, N. D., Rogers, B., & Heckers, S. (2011). Functional resting-state networks are differentially affected in Schizophrenia. *Schizophrenia research 130*(1–3), 86–93.

Woolrich, M. W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., Beckmann, C., Jenkinson, M., & Smith, S. M. (2009). Bayesian analysis of neuroimaging data in FSL. *NeuroImage 45*, S173–S186.

Xia, M., Wang, Z., Dai, Z., Liang, X., Song, H., Shu, N., Li, K., & Y, H. (2014). Differentially disrupted functional connectivity in posteromedial cortical subregions in Alzheimer's disease. *J Alzheimers Dis 39*(3), 527–43.

Yang, X., Yao, C., Tian, T., Li, X., Yan, H., Wu, J., Li, H., Pei, L., Liu, D., Tian, Q., Zhu, L.-Q., & Lu, Y. (2016). A novel mechanism of memory loss in Alzheimer's disease mice via the degeneration of entorhinal–CA1 synapses. *Molecular Psychiatry 23*, 199–210.

Yu, Q., Allen, E. A., Sui, J., Arbabshirani, M. R., Pearlson, G., & Calhoun, V. D. (2016). Brain connectivity networks in Schizophrenia underlying resting state functional magnetic resonance imaging. *Current Topics in Medicinal Chemistry 12*(21), 2415–2425.

Yun, S. H., Gamkrelidze, G., Stine, W. B., Sullivan, P. M., Pasternak, J. F., LaDu, M. J., & Trommer, B. L. (2006). Amyloid-beta(1–42) reduces neuronal excitability in mouse dentate gyrus. *Neurosci Lett 403*(0), 162–165.

Zhan, S. S., Beyreuther, K., & Schmitt, H. P. (1993). Quantitative assessment of the synaptophysin immuno-reactivity of the cortical neuropil in various neurodegenerative disorders with dementia. *Dementia 4*(3), 66–74.

Zhang, B., & Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology 4*.

Zhang, L., Guindani, M., Versace, F., Engelmann, J. M., & Vannucci, M. (2016). A spatiotemporal nonparametric Bayesian model of multi-subject fMRI data. *Ann. Appl. Stat. 10*(2), 638–666.

Zhang, L., Guindani, M., Versace, F., & Vannucci, M. (2014). A spatio-temporal nonparametric Bayesian variable selection model of fMRI data for clustering correlated time courses. *NeuroImage 95*, 162–175.

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging 20*(1), 45–57.

Zhao, X., Lui, Y., Wang, X., Liu, B., Xi, Q., Guo, Q., Jiang, H., Jiang, T., & Wang, P. (2012). Disrupted small-world brain networks in moderate Alzheimer's disease: a resting-state fMRI study. *PloS One 7*(3), :e33540.

Zhou, S., & Yu, Y. (2018). Synaptic E–I balance underlies efficient neural coding. *Front Neurosci 12*, 46.

Zilberter, M., Ivanov, A., Ziyatdinova, S., Mukhtarov, M., Malkov, A., Alpár, A., Tortoriello, G., Botting, C. H., Fülöp, L., Osypov, A. A., Pitkänen, A., Tanila, H., Harkany, T., & Zilberter, Y. (2013). Dietary energy substrates reverse early neuronal hyperactivity in a mouse model of Alzheimer's disease. *J. Neurochem. 125*(1), 157–171.

Band / Volume 188
**Finite-Difference Time-Domain Simulations Assisting to Reconstruct the Brain's Nerve Fiber Architecture by 3D Polarized Light Imaging**
M. Menzel (2018), ix, 296 pp
ISBN: 978-3-95806-368-6

Band / Volume 189
**Characterization of the cell-substrate interface using surface plasmon resonance microscopy**
E. M. Kreysing (2018), xiii, 260 pp
ISBN: 978-3-95806-369-3

Band / Volume 190
**Scattering! Soft, Functional and Quantum Materials**
Lecture Notes of the 50th IFF Spring School 2019
11 – 22 March 2019, Jülich, Germany
ed. by M. Angst, T. Brückel, S. Förster, K. Friese, R. Zorn (2019),
ca 1000 pp
ISBN: 978-3-95806-380-8

Band / Volume 191
**Absolute scale off-axis electron holography of thin dichalcogenide crystals at atomic resolution**
F. Winkler (2019), xxiii, 187 pp
ISBN: 978-3-95806-383-9

Band / Volume 192
**High-resolution genome and transcriptome analysis of *Gluconobacter oxydans* 621H and growth-improved strains by next-generation sequencing**
A. Kranz (2019), III, 182 pp
ISBN: 978-3-95806-385-3

Band / Volume 193
**Group IV (Si)GeSn Light Emission and Lasing Studies**
D. Stange (2019), vi, 151 pp
ISBN: 978-3-95806-389-1

Band / Volume 194
**Construction and analysis of a spatially organized cortical network model**
J. Senk (2019), 245 pp
ISBN: 978-3-95806-390-7

Schriften des Forschungszentrums Jülich
Reihe Schlüsseltechnologien / Key Technologies

Band / Volume 195
**Large-scale Investigations of Non-trivial Magnetic Textures in Chiral Magnets with Density Functional Theory**
M. Bornemann (2019), 143 pp
ISBN: 978-3-95806-394-5

Band / Volume 196
**Neutron scattering**
Experimental Manuals of the JCNS Laboratory Course held at Forschungszentrum Jülich and at the Heinz-Maier-Leibnitz Zentrum Garching
edited by T. Brückel, S. Förster, G. Roth, and R. Zorn (2019),
ca 150 pp
ISBN: 978-3-95806-406-5

Band / Volume 197
**Topological transport in non-Abelian spin textures from first principles**
P. M. Buhl (2019), vii, 158 pp
ISBN: 978-3-95806-408-9

Band / Volume 198
**Shortcut to the carbon-efficient microbial production of chemical building blocks from lignocellulose-derived D-xylose**
C. Brüsseler (2019), X, 62 pp
ISBN: 978-3-95806-409-6

Band / Volume 199
**Regulation and assembly of the cytochrome *bc*$_1$-*aa*$_3$ supercomplex in *Corynebacterium glutamicum***
C.-F. Davoudi (2019), 135 pp
ISBN: 978-3-95806-416-4

Band / Volume 200
**Variability and compensation in Alzheimer's disease across different neuronal network scales**
C. Bachmann (2019), xvi, 165 pp
ISBN: 978-3-95806-420-1

Weitere ***Schriften des Verlags im Forschungszentrum Jülich*** unter
http://wwwzb1.fz-juelich.de/verlagextern1/index.asp

Schlüsseltechnologien / Key Technologies
Band / Volume 200
ISBN 978-3-95806-420-1

Mitglied der Helmholtz-Gemeinschaft

JÜLICH
Forschungszentrum