Investigating connectivity in brain-like networks

The University Of Sheffield.

Umberto Esposito

Department of Computer Science
Faculty of Engineering

This dissertation is submitted for the degree of
Doctor of Philosophy

Supervisor: Dr. Eleni Vasilaki

February 2016

Dedico questa tesi a mamma e papa', che, nonostante le tante difficolta', hanno sempre trovato il modo di supportarmi durante tutto il percorso di studi, che ora giunge a compimento. Grazie

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and are the result of my own work and research during the three years of my PhD. Most of the material presented has already been peer-reviewed and published, or is about to be, in the form of journal papers with myself as first author. None of the contents has been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university.

Umberto Esposito
February 2016

Acknowledgements

Special thanks must go to my supervisor, Eleni Vasilaki, for her wise guidance throughout the entire PhD, countless advices and ideas without which this work would not have been possible. Most of all, thanks for the trust showed from the very beginning.

Thanks to my parents, for giving me continuous and unconditional support even from thousands of kilometers away; to my three elder brothers Giacomo, Gianluca and Alessandro, and the three sisters in law Alessandra, Daniela and Livia, for being always close and present; to my nephew Leonardo and to my nieces Sara and Arianna, who, even without understanding the meaning of all this, always bring sparks of happiness and lightness. Also, thanks to my favourite uncle and aunt Tonino and Pina, for being present and showing constant interest.

Dulcis in fundo, I would like to acknowledge very much Gabriela, for her love and priceless support, especially during these months while writing the thesis. Thanks for the many moments of joy and for unconditionally believing in me, challenging me and making me improve myself greatly.

Abstract

Experimental research over the last two decades has shown that the anatomical connectivity among neurons is largely non-random across brain areas. This complex organisation shapes the flow of information, giving rise to specific pathways and motifs, which are ultimately responsible for processes like emotions, cognitive functions and behaviour, just to mention few. Due to the spectacular progress of technology, the study of the brain wiring diagram, known as *connectomics*, has received considerable attention in recent years, resulting in the proliferation of large data sets. From one side, this adds a significant contribution towards a better understanding of the complex processes that take place in the brain. On the other side, however, analysing such large connectivities is a hard task that has not yet found a satisfactory solution. Particular evidence has been found for bidirectional motifs, occurring when two neurons project onto each other via connections of equal strength, and unidirectional motifs, when one of the two connections is dominant. These specific motifs were found to correlate with short-term synaptic plasticity properties, which are related to resources availability for signal transmission. The aim of this thesis is to add a contribution to the ongoing efforts spent on answering the two main questions related to motif evidence: *How* can we satisfactory detect and measure motifs in large networks and *why* do they have the characteristics that we observe? Following existing literature, we hypothesise that bidirectional and unidirectional motifs appear as a consequence of learning processes, which move the distribution of the synaptic connections away from randomness through activity dependent synaptic plasticity. Based on this, we introduce a symmetry measure for global connectivity and a statistics-based heuristic algorithm for directed and weighted graphs that is able to detect overlapping bidirectional communities within large networks. On the other side, to address the *why* question we introduce an error-driven learning framework for short-term plasticity that acts jointly with Spike-Timing Dependent Plasticity, a well-known learning mechanism for long-term plasticity: By allowing synapses to change their properties, neurons are able to adapt their own activity depending on an error signal. This results in more rich dynamics and also, provided that the learning mechanism is target-specific, leads to specialised groups of synapses projecting onto functionally different targets, qualitatively replicating the experimental results of Wang and collaborators in 2006.

Table of contents

Chapter 1

Introduction and Motivation

Every moment during our life we are exposed to an enormous quantity of stimuli flowing through the sensory apparatus to eventually reach the brain. Here, incoming signals can trigger all sorts of processes: From learning an action to memorising a name, from solving an equation to producing the feeling of joy. Despite the size, it is outstanding the level of complexity and organization that the brain of an animal can achieve and its capability to fulfil a great quantity of different tasks in a very short interval of time.

Certainly, the number of neurons plays an important role in determining the level of complexity of brains and individuals: An adult human brain typically consists of $\sim 10^{11}$ neurons, many more than cats, rats and octopuses for instance. However, the fact that an elephants' brain has more than double the neurons of a human brain suggests that it is not only about the quantity. What really determines brain's properties it is the way neurons are connected to each other: It is a recent dominant belief, indeed, that cognitive functions are stored in the so-called *connectome*, the wiring diagram of the brain (Lichtman et al. [2008]; Sporns et al. [2005]). *"We are our connectome"*, says Sebastian Seung, one of the pioneers of these theory (Seung [2009]). During the last decades, the investigation of the brain wiring diagram, known as *connectomics*, has generated great excitement (Lichtman and Sanes [2008]; Seung [2009]; Van Essen et al. [2012]) as the comprehension of this level of organisation (Kandell et al. [2008]) is thought to be pivotal to understanding the richness of high-level cognitive, computational and adaptive properties of the brain, as well as its dysfunctions.

Such excitement has shaped the recent research history in Neuroscience. Novel discoveries in molecular biology (Lichtman et al. [2008]; Wickersham et al. [2007]; Zhang et al. [2007]), neuroanatomical methods (Chklovskii et al. [2010]; Denk and Horstmann [2004]), electrophysiology (Hai et al. [2010]; Perin et al. [2011]; Song et al. [2005]) and imaging (Friston [2011]; Minderer et al. [2012]; Wedeen et al. [2012]) have pushed forward

the technological limits for ultimate access to neuronal connectivity. As a consequence, also due to improvements in technology, experimental techniques and computational paradigms (Luo et al. [2008]; Smith [2007]), connectomics has made significant progress (Bock et al. [2011]; Briggman et al. [2011]; Varshney et al. [2011]; White et al. [1986]), resulting in a rapid proliferation of neuroscience datasets (Briggman and Denk [2006]; Insel et al. [2003]; Koslow and Subramaniam [2005]; Kötter [2001]).

The most evident proof of the ferment generated by this new discipline and the possible future impact of its results can be found in the fact that an unprecedented amount of funding has been destined to two highly ambitious and recently launched projects, respectively in Europe and United States: The Human Connectome Project[1] – mapping the human connectome as accurately as possible in a large number of normal adults (Toga et al. [2012]) – and The Human Brain Project[2] – aiming to simulate the human brain with supercomputers (Frackowiak and Markram [2015]). A complete map of the connections between neurons could provide an unprecedented and extremely powerful knowledge, with great benefits, for instance, in diseases treatment (Van Essen and Ugurbil [2012]; Wang et al. [2013]; Zhou et al. [2012]).

Connectomics is a field that develops within the older *Connectionism Theory* (Rumelhart et al. [1986]; Sejnowski and Rosenberg [1987]; Thorndike [1911]): Neurons involved in the same functions are grouped together to form clusters working as an unique elementary unit. The interaction of these clusters determines brain's behaviour. In this picture, connections between neurons cannot be randomly distributed, but rather they are expected to form highly specialized patterns, reflecting the functionality of the clusters and their intrinsic features. This view is now dominant and has received substantial validation from experiments: Several works reported a significant excess of particular connectivity motifs in different areas of the brain (Lefort et al. [2009]; Perin et al. [2011]; Silberberg and Markram [2007]; Song et al. [2005]; Wang et al. [2006]), suggesting that indeed connectivity is generally not random (Sporns [2011b]).

The process through which such a specialisation of connections takes place is called learning, which, at a a neuronal level, allows the formation of patterns within the connectivity (Bienenstock et al. [1982]; Hebb [1949]; Song and Abbott [2001]). This is possible thanks to synaptic plasticity: The property of synapses to change their efficacy over time according to local activity and available resources (Thomson [2000]), resulting in a modulation of information flow (Abbott and Nelson [2000]; Abbott and Regehr [2004]; Fuhrmann et al. [2002]). Synaptic plasticity processes are usually divided into short-term and long-term,

[1] http://www.humanconnectomeproject.org
[2] https://www.humanbrainproject.eu

which differ not just in the duration of the synaptic modification but also in the underlying biological mechanism (Fioravante and Regehr [2011]; Kullmann and Lamsa [2007]; Regehr [2012]) and functional role (Billings and van Rossum [2009]; Gütig and Sompolinsky [2006]; Natschläger et al. [2001]). Both forms of plasticity can lead to a strengthening or a weakening of the synapse.

The occurrence of stereotypical connectivity motifs as mentioned above has been accompanied in some works by physiological information on activity-dependent short-term and long-term plasticity (Buonomano and Merzenich [1998]) and rewiring (Chklovskii et al. [2004]; Le Be' and Markram [2006]). Moreover, different motifs seem to correlate with different synaptic properties (Pignatelli [2009]; Wang et al. [2006]), which in turn are related to signal transmission, underlying learning mechanisms and eventually cognitive functions and behaviour (Bressler and Menon [2010]; Bullmore and Sporns [2009]; Lichtman et al. [2008]): Neurons connected by synapses exhibiting short-term facilitation seem to form mostly reciprocal (bidirectional) motifs, whereas neurons connected by synapses exhibiting short-term depression seem to form unidirectional motifs.

This great amount of data that has been collected during past years has certainly contributed to shed light on many aspects about connectivity in brain-like circuits, but the two big questions still remain unanswered: How and when can we confidently reveal and measure regions of networks that show non random properties in the connectivity? And why do such motifs form and correlate with specific synaptic properties? The research I present here addresses both aspects, and similarly to Clopath et al. [2010] and Vasilaki and Giugliano [2012, 2014] we hypothesise that motifs are regions that have been shaped via activity dependent synaptic plasticity processes, and therefore that learning moves the distribution of the synaptic connections away from the initial random condition that would correspond to unspecialised synapses.

The first step towards a correct interpretation of the huge amount of data coming from brain networks is revealing their structural and functional properties. To achieve this, principles and tools from Graph and Network Theory have been applied to brain networks (He and Evans [2010]; Sporns [2011a, 2013]) with promising results (Bassett and Bullmore [2009]; Guye et al. [2010]). In general, several studies have demonstrated that many real world processes can be modelled in terms of complex networks (Albert and Barabási [2002]; Barabási and Oltvai [2004]; Green and Sadedin [2005]; Newman [2010]), making the study of networks' topology and properties a topic of major interest within the entire scientific community. Of particular relevance for brain networks is the problem of structures' detection as it is directly related to motifs formation. Typical Graph Theory problems dealing with structure searching, for instance sub-graph isomorphism and clique identification (Bondy and

Murty [2008]; Diestel [2010]), are proven to be either NP-complete or NP-hard (Bomze et al. [1999]; Cook [1971]; Garey and Johnson [1990]; Papadimitriou [1977]; Wegener [2005]). Extensive search is therefore impracticable and feasible approaches are based on heuristic search or on algorithms looking for sub-optimal solutions. Even with these approaches, the computational complexity grows very quickly and explodes for just few thousands of nodes (Fortunato [2010]), hence making impossible to perform an effective and accurate search on large networks within a relatively small time scale.

This thesis is built around the three journal papers that I produced as first author during my PhD [3]:

I **Esposito U, Giugliano M, van Rossum M, Vasilaki E** (2014) *Measuring Symmetry, Asymmetry and Randomness in Neural Network Connectivity.* PLoS ONE 9(7): e100805.

II **Esposito U, Giugliano M and Vasilaki E** (2015) *Adaptation of short-term plasticity parameters via error-driven learning may explain the correlation between activity-dependent synaptic properties, connectivity motifs and target specificity.* Front. Comput. Neurosci. 8:175.

III **Esposito U and Vasilaki E** (submitted) *Detection of multiple and overlapping bidirectional communities within large, directed and weighted networks of neurons.* Nature Sci. Rep..

After a brief literature review in **Chapter 2**, each of the three following chapters is devoted to one article, with **Chapter 3** and **Chapter 4** addressing the problem of motifs detection within networks of neurons and **Chapter 5** focussed on the experimental correspondence between motifs and synaptic properties. Specifically, in Chapter 3 a statistical measure is introduced, which responds to the problem of determining the symmetry of the global connectivity of a network of neurons. Following this, in Chapter 4 a heuristic algorithm is presented, designed to identify a particular class of local structures within such networks. In Chapter 5 a learning scheme for short-term plasticity is defined, which, combined with a traditional learning method for long-term plasticity, is able to reproduce the correspondence between motifs and short-term synaptic properties. Finally, **Chapter 6** is devoted to a discussion about limitations of the present work and future improvements. Note that almost the entire content of the papers has been reported in Chapters 3,4,5, with minor paragraphs that have been moved to Introduction and Chapter 6 to obtain overall a better flow for the benefit of the reader.

[3] At the moment of writing, two published and one submitted

Chapter 2

Literature Review

The most complex part of the human body is certainly the brain, not only for its anatomical complexity but also for its puzzling functioning. It is therefore not surprising that understanding brain's functions has attracted the curiosity of generations of people since the dawn of human civilisation: First evidence of neuroscience practices dates back to ancient Egyptians, Greeks and Romans, for example the mummification procedure. Available equipments at those times were very primitive, resulting in a vague, lacking and sometimes even erroneous knowledge. For instance, it was a common belief that the heart is the seat of intelligence, reason why the brain was regularly removed in preparation for mummification whereas the heart was the only organ left. Residual of this ancient belief is for instance the expression "knowing by heart".

Since then, outstanding progresses have been made, starting from these ancient civilisations themselves. Notably, the Greek *Hippocrates of Kos* (*460 - 370 BC*) made exceptional contributions to Medicine and he is still considered the father of Western Medicine. The Egyptian *Herophilius of Chalcedon* (*335/330 - 280/250 BC*) researched the anatomy and physiology of the nervous system and was able to distinguish between cerebrum and cerebellum. The Greek physician and surgeon *Galen* (*129 - 200 AD*), who lived under the Roman Empire, dissected brains of several non-human mammals and recognised the function of the cerebellum as controlling muscles, separated from the cerebrum, thought to regulate the senses. Furthermore, he noted that specific spinal nerves control specific muscles, which remained the most advanced understanding of spinal functions until the 18^{th} and 19^{th} century.

Indeed, we had to wait until three centuries ago to find significant advances in brain's studies, when microscope started to be widespread. Major contributions came from two Italians: *Luigi Galvani* (*1737 - 1798*), who first observed the role of electricity in nerves, and *Camilo Golgi* (*1843 - 1926*), who developed a staining procedure to reveal structures of neurons. By using and improving this technique, the Spanish *Santiago Ramón y Cajal* (*1852 -*

1934) provided descriptions of neurons in the central nervous system and produced excellent depictions of structures and their connectivity which are still in use. *"In recognition of their work on the structure of the nervous system"*, in 1906 Golgi and Ramón y Cajal were awarded with the Nobel Prize in Physiology or Medicine. Ramón y Cajal is considered the father of modern Neuroscience for his original pioneering investigation and for experimentally demonstrating that relationships between nerves are not continuous but contiguous. This finding provided a strong evidence in favour of the neuron doctrine, of which Ramon y Cajal was a promoter, in opposition to the reticular theory supported by Golgi (López-Muñoz et al. [2006]). The neuron theory originated in 1891 with the introduction of the term neuron by the German neuroanatomist *Heinrich von Waldeyer-Hartz (1836 - 1921)* (Ramón y Cajal [1954]) and found the definite success only in the 1950s with the development of electron microscopy (Markram et al. [2011]). Thanks to these studies in the first part of the 20^{th} century, Neuroscience started to be recognised as a distinct discipline rather than being part of Medicine. From the second half of the 20^{th} century, due to the phenomenal advances in technology that have been continuously improving the experimental techniques, the study of the nervous system has been receiving a formidable boost, allowing scientists to study the brain in all its aspects.

The most important function that a brain can perform is probably *Leaning*: It allows to speak, to survive, to not fall from the bicycle. Everyday we learn from new stimuli, sometimes without even being conscious of it: Our brain works for us, continuously acquiring and processing information, in a way that is not just largely out of our control but it is also way beyond our current understanding. What we observe in daily life is the macroscopic manifestation of learning, at a behavioural level, as a result of the interaction of many processes within the brain. Indeed, on a microscopic scale, learning is the interaction between neurons whose activity triggers complex chains of proteins, resulting in activation of different brain areas. Most of the structures involved are part of the cerebral cortex: From the visual cortex, and in general sensory cortex, to the hippocampus - required for the formation of long-term memories and implicated in the maintenance of cognitive maps for navigation - and amygdala - involved in signaling the cortex of motivationally significant stimuli such as those related to reward.

The cerebral cortex is a sheet of neural tissue, $2-4\,mm$ thick in humans, that covers cerebrum and cerebellum in the mammalian brain. Besides various types of neurons, it contains also a large number of 'supporter' cells, so-called glia cells, that are required for energy supply and structural stabilization of brain tissue. Since glia cells are not directly involved in information processing, they will not be discussed any further. Furthermore, throughout this thesis only spiking neurons of the cortex will be used, hence neglecting other

few rare types, such as for instance analog neurons in the mammalian retina (Gerstner and Kistler [2002b]).

2.1 Biological neuron

Neurons are electrically excitable cells that are at the basis of the nervous system. Similarly to other cells, they are equipped with a membrane, that consists of a thin bilayer of lipids separating and protecting the interior of the neuron from the outside. The cell membrane, however, is not a perfect electrical insulator: Specific proteins are embedded in it acting as ion gates, through which ions can move from one side to the other. Ions gates can be of two types: Ions channels, which transport ions passively, i.e. through diffusion, and ions pumps, doing it actively to maintain a voltage difference across the membrane itself. Also, each ions gate can be specialised for only one type of ion. The most common ions are Na^+, K^- and Ca^{++} (Gerstner and Kistler [2002b]).

Action potential From a functional point of view, a typical neuron can be divided into three parts: Dendrites, axon and soma, respectively input, output and central processing units. Incoming signals travel along the dendrites towards the soma, which performs the non-linear processing at the basis of the spiking mechanism: If the total input exceeds a certain threshold then the output signal is generated and delivered to other neurons by travelling through the axon. The signal is an electrical pulse, called action potential or spike. Neurons that are repeatedly stimulated can elicit sequences of spikes one shortly after another, which are called spike trains, but single action potentials are still usually well separated from each other: No matter the strength of the input, it is normally impossible to excite a second spike during or immediately after a first one. This short time interval is called refractory period (Gerstner and Kistler [2002b]).

2.1.1 Dynamic properties

The state of a neuron can be described by its membrane potential $u(t)$, defined as the voltage difference between the interior of the cell and its surroundings. The different states during the action potential generation process are briefly described below and are depicted in Fig. 2.1 for an excitatory neuron by means of a schematic drawing.

Resting potential Without any input there is no activity in the postsynaptic neuron and its membrane potential is at its (constant) resting value, $u(t) = u_{rest}$. In this situation the cell

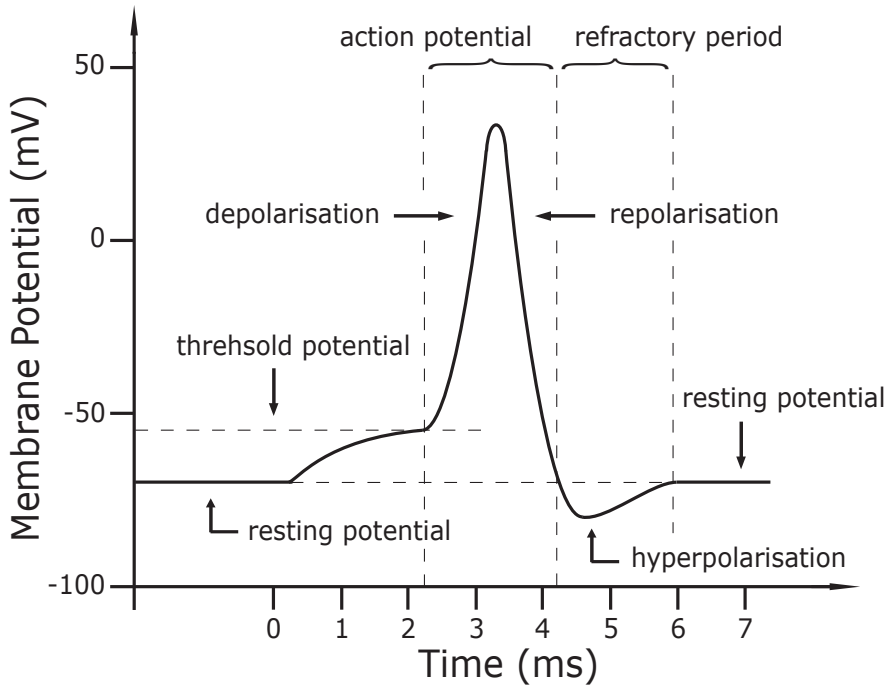

Figure 2.1 Schematic behaviour of the membrane potential of an excitatory neuron before, after and during an action potential. Five phases can be distinguished: *1.* The neuron is not receiving any input and the membrane potential is at its resting level. *2.* Incoming signals starts to excite the neuron, each of them generating an excitatory postsynaptic potential (EPSP) which sum up in a way that is approximately linear (Gerstner and Kistler [2002b]). *3.* Once the threshold is reached, the membrane potential suddenly jumps and goes back towards the resting level, describing the characteristic shape of the action potential. *4.* Once it reaches the resting value, the membrane potential continues to decrease giving rise to the refractory period. *5.* Finally, the potential goes back to the resting value, ready for a new cycle [figure produced with Adobe Illustrator®].

membrane has already a strong negative polarization of about $-65\,mV$, so as the equilibrium is actually a dynamical equilibrium in which ion pumps compensate the natural diffusion of the ions from one side to the other of the neuron membrane through the ion channels (Gerstner and Kistler [2002b]).

Activation Whenever a presynaptic neuron releases a spike, this is transmitted to the postsynaptic neuron i and generates a postsynaptic potential (PSP) whose amplitude is usually of about $1\,mV$ that alters the membrane potential of the postsynaptic neuron u_i, in two possible ways. If the change is positive, i.e. $u_i(t) - u_{i,rest} > 0$, then the postsynaptic potential is excitatory (EPSP), the change in the membrane potential is called depolarisation and the synapse connecting the emitting neuron and neuron i is said to be excitatory. Otherwise,

if $u_i(t) - u_{i,rest} < 0$, then the postsynaptic potential is inhibitory (IPSP), the change in the membrane potential is called hyperpolarization and the synapse is said to be inhibitory (Gerstner and Kistler [2002b]). In excitatory neurons, an incoming signal from a presynaptic neuron leads to the opening of Sodium channels and the subsequent influx of Na^+ into the cell is what causes the EPSP. Excitatory synapses are largely the most studied type of synapse, but now is becoming clear that inhibitory synapses play a crucial and indispensable role in stabilizing the activity of networks of neurons (Brunel [2000]; Silberberg and Markram [2007]; Vogels and Abbott [2009]; Vogels et al. [2011]).

Spike generation As long as there are only few input signals, the total change in the membrane potential is approximately the sum of the individual PSPs. If too many input spikes arrive during a short interval then the membrane potential reaches a critical threshold, which generally is $\sim 20-30\,mV$ above the resting potential, causing the opening of even more Sodium channels. This results in a consistent influx of Na^+ ions which gives rise to a sudden depolarisation up to $\sim +30\,mV$ (see Fig. 2.1). At this value of the potential, Sodium channels starts to close and Potassium channels start to open. Differently from Sodium, K^+ ions diffuse from the interior of the cell to the extracellular medium, bringing the potential back towards its resting value. A typical pulse has a total amplitude of about $100\,mV$ and a duration of $1-2\,ms$ (Gerstner and Kistler [2002b]).

Refractoriness After the pulse, the membrane potential passes through a phase of hyperpolarization below the resting value, which identifies the absolute and relative refractory periods, and slowly reaches the resting value. Neurons can elicit sequences of spikes one shortly after another, which are called spike trains, but single action potentials are still usually well separated from each other: No matter the strength of the input, it is normally impossible to excite a second spike during or immediately after a first one. The minimal time interval between two spikes is normally $\sim 1\,ms$ (Gerstner and Kistler [2002b]). After hyperpolarisation, Sodium and Potassium pumps eventually bring the membrane bacl to the resting vlaue (Gerstner and Kistler [2002b]).

2.1.2 Synaptic transmission

The site where the action potential is transmitted from the sending neuron, called presynaptic neuron, to the receiving cell, the postsynaptic neuron, is called synapse. The most common type of synapse in the vertebrate brain is the chemical synapse. Here, the terminal part of the presynaptic axon comes very close to the dendrites of the postsynaptic cell, leaving only a tiny gap between them, called synaptic cleft. The signal transmission at a chemical

synapse can be simplified as a two-step process: The electrical signal from the presynaptic neuron is translated into a chemical one that travels through the synaptic cleft and eventually is re-translated into an electrical stimulus propagating within the postsynaptic cell (Dayan and Abbott [2001]).

This mechanism is possible because of the structural properties of a neuron: The axon terminal, also called bouton, contains synaptic vesicles which in turn accommodate a variety of chemical substances called neurotransmitters. These endogenous chemicals are the entities travelling across the synapse, hence permitting the transmission of the signal. The entire mechanism is triggered by the arrival of an action potential in the presynaptic bouton, which causes a depolarisation that opens the voltage-dependent calcium channels on the presynaptic neuron. Due to diffusion, Ca^{++} ions enter the presynaptic cell and drive the vesicles towards the membrane. Here, vesicles merge with the membrane causing the release of the neurotransmitters into the synaptic cleft (exocytosis). Neurotransmitters then diffuse towards the postsynaptic site and here they bind with the receptors located on the membrane. This leads to the opening of the ion channels on the postsynaptic neuron, so that ions from the extracellular fluid flow into the cell. The ion influx, in turn, leads to a change of the membrane potential at the postsynaptic site, completing the transmission of the signal. The voltage response of the postsynaptic neuron to a presynaptic action potential is called postsynaptic potential (PSP) which can be excitatory (EPSP) or inhibitory (IPSP), depending on the nature of the neuron. (Dayan and Abbott [2001]).

Receptors and neurotransmitters Receptors on the postsynaptic site can be of two types, ionotropic or metabotropic. The first class directly activates the ion channel, whereas the second class does it indirectly by means of an intracellular messenger. Hence, ionotropic processes are faster than metabotropic, and, in addition, metabotropic receptors can induce long-lasting changes inside a neuron. Dopamine and serotonin are instances of neuromodulators that act through metabotropic receptors. At the level of neurotransmitters, the most common ones in the brain are glutamate and GABA (γ-aminobutyric acid), respectively excitatory and inhibitory, which can act both ionotropically and metabotropically. The principal ionotropic receptors for glutamate are AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor) and NMDA (N-Methyl-D-aspartic acid). Activation and deactivation of AMPA receptors is faster if compared to NMDA (Dayan and Abbott [2001]).

NMDA and synaptic strength NMDA receptors are particularly important because they show a dependence on the postsynaptic potential that is not normally displayed by other receptors. This voltage dependence is due to the fact that when the postsynaptic neuron is

close to its resting state, Mg^{++} ions block the NMDA receptors themselves. To remove the blockage, and therefore to activate the ion channel, the postsynaptic neuron must be depolarised. Although the channels controlled by NMDA are permeable to sodium and potassium, their permeability to calcium is much greater. Interestingly, Ca^{++} ions are of crucial importance for long-term modification of synaptic strength (Dayan and Abbott [2001]).

2.2 Models of neuron

2.2.1 Hodgkin-Huxley biological model

According with the pioneering work of Hodgkin and Huxley (Hodgkin and Huxley [1952]) on the giant axon of the squid, the neuron dynamics can be described in terms of an *RC*-circuit, with the membrane being the capacitor and the ions channels being the resistances. Therefore, if there is an external current $I(t)$ – it can be either an injected current or a PSP or both – it may add further charge to C or leak through the ion channels. Therefore we can write the typical equation describing the *RC*-circuit (Gerstner and Kistler [2002b]):

$$C\frac{du}{dt} = -\sum_k I_k(t) + I(t) \tag{2.1}$$

Where the sum runs over all the ionic currents passing through the cell membrane. In the standard Hodgkin-Huxley model there are only three types of channels: Sodium and potassium channels, modelled through time and voltage-dependent conductances g_{Na} and g_K, and an unspecific leakage channel with fixed conductance $g_L = 1/R$. If all channels are open, they transmit currents with a maximum conductance g_{Na} or g_K, respectively. Normally, however, some of the channels are blocked. The probability that a channel is open is described by additional variables m, n, and h. The combined action of m and h controls the Na channels, whereas the K gates are controlled by n .

The equations of Hodgkin and Huxley provide a good description of the electrophysiological properties of the giant axon of the squid. These equations capture the essence of spike generation by sodium and potassium ion channels and this mechanism is essentially preserved in higher organisms.

2.2.2 Formal spiking neuron models

Hodgkin-Huxley model is the basis for many other more complicated and realistic models. However, detailed conductance-based neuron models like these can reproduce electrophysiological measurements to a high degree of accuracy, but because of their intrinsic complexity they are difficult to analyse. For this reason, simple phenomenological spiking neuron models are highly popular for studies of neural coding, memory, and network dynamics.

Leaky integrate-and-fire One of the most popular neuron models is the leaky integrate-and-fire model (LIF) (Gerstner and Kistler [2002b]). The neuron is still described by the *RC* equation, however the way the action potentials are modelled - as delta of Dirac - is much more simple. If we multiply by R and introduce the time constant $\tau_m = RC$ for the leaky integrator in Eq. (2.1), we obtain the standard form:

$$\tau_m \frac{du(t)}{dt} = -u(t) + RI(t) \tag{2.2}$$

The combination of this leaky integration and the firing and reset condition

$$\lim_{t \to t^{(f)+}} u(t) = u_r \quad \text{for} \quad t^{(f)} \,|\, u(t^{(f)}) = \vartheta \quad \text{and} \quad \left.\frac{du(t)}{dt}\right|_{t=t^{(f)}} > 0 \tag{2.3}$$

defines the leaky integrate-and-fire model. A more complex model is the Spike Response Models (SRM), where one can take into account the time evolution of the action potential, of the PSPs and also of the refractory period (Gerstner and Kistler [2002b]).

Modelling synapses The above equations specify the neuron but not the synapse. Indeed, the current in the equation can be an external injected current I_{ext} or a stimulus from other neurons, which before arriving in the soma, passes through the synapse. In the simplest case we can model these currents as delta pulses multiplied by the strength of the synapse:

$$I_{syn} = \sum_j w_{ij} \sum_f \delta(t - t_j^f) \tag{2.4}$$

In a more complex case we can consider an exponential time dependence instead of the delta. Even more complex is the case of conductance-based synapses, as used in the Hodgkin-Huxley model (Gerstner and Kistler [2002b]).

2.3 Synaptic plasticity

It is widely believed that learning processes are related to modifications of synaptic connections between neurons in the brain. The ability of a synapse to strengthen or weaken over time is called synaptic plasticity and it has been theoretically postulated and subsequently experimentally confirmed.

Well before the discovery of the existence of synapses, the Scottish philosopher *Alexander Bain (1818 – 1903)* was probably the first one to introduce the concept of junctions between neurons (Markram et al. [2011]), postulating that changes of these junctions are at the basis of learning and memories (Bain [1873]). Few years later, the American philosopher and psychologist *William James (1842 – 1910)* speculated on the cause that would strengthen the junctions, introducing the idea of learning through correlation and linking plasticity with behavioural habits (James [1980]). The very first hypothesis that associative memories and practice-dependent motor skills may depend on a localized facilitation of synaptic transmission (Berlucchi and Buchtel [2009]) was put forward by the Italian neuropsychiatrist *Eugenio Tanzi (1856 – 1934)* in 1893 (Tanzi [1893]), 4 years before the coming into existence of the term synapse (Tansey [1997]), due to the English neurophysiologist *Sir Charles Scott Sherrington (1857 - 1952)* (Sherrington [1897]) [1]. Few years later, the Italian psychiatrist *Ernesto Lugaro (1870 - 1940*) expanded on Tanzi's work by explicitly connecting for the first time the concepts of plasticity and plastic activity, already in use at that time, with his teacher's hypothesis of functional modifiability of synapses (Lugaro [1913]), a denotation that persists to this day (Berlucchi and Buchtel [2009]; Markram et al. [2011]). Both Tanzi and Lugaro were fervent admirers of Cajal's neuron theory and Cajal himself complemented Tanzi's hypothesis with his own view of plasticity as the result of the formation of new connections between cortical neurons (Berlucchi and Buchtel [2009]).

With Cajal, in the first two decades of the 20^{th} century, it was generally accepted that learning is based on a reduced resistance at exercised synapses, and that neural processes become associated by coactivation. Subsequently, due to the appearance of other conjectures evidently more appealing, this point of view lost recognition and it was only in the late 1940s that the synaptic plasticity theory of learning was rehabilitated when the Polish neurophysiologist *Jerzy Konorski (1903 - 1973)* and particularly the Canadian psychologist *Donald Olding Hebb* (*1904 - 1985*) argued successfully that there was no better alternative way to think about the modifiability of the brain by experience and practice (Berlucchi and Buchtel [2009]; Hebb [1949]; Konorski [1948]).

[1] Sherrington received the Nobel Prize in Physiology or Medicine with Edgar Adrian, 1st Baron Adrian, in 1932. They showed that reflexes require integrated activation and demonstrated reciprocal innervation of muscles (Sherrington's law).

Hebbian rule Hebb has the credit of having rephrased the idea that the efficacy of a synapse can change due to correlated neuronal activity: *When an axon of cell j is near enough to excite cell i or repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that j's efficiency, as one of the cells firing i, is increased (Hebb [1949]).* Despite Hebb himself claimed that he was *"not proposing anything new"* (Berlucchi and Buchtel [2009]), his formulation directly inspired thousands of studies and strengthening of connections between co-active cells has become known as Hebbian plasticity or Hebbian learning rule.

The first experimental evidence confirming Hebb's postulate was published two decades after: Tetanic stimulation (high-frequency stimulation) on the dentate gyrus of rabbit hippocampus was shown to be responsible for a persistent growth of the postsynaptic response (Bliss and Lømo [1970, 1973]; Lømo [1964, 2003]). This phenomenon, initially called long-lasting potentiation (Bliss and Lømo [1973]), is now known as long-term potentiation (LTP) (Douglas and Goddard [1975]) and it was the first form of synaptic plasticity to be studied. Since then, different mechanisms of synaptic plasticity have been discovered, and nowadays two main types can be distinguished: Long-term and short-term plasticity.

Long-term plasticity and in particular Hebbian learning and Spike-Timing-Dependent-Plasticity (see below) are forms of unsupervised learning, which captures correlations in the neuronal input. Hence, their involvement in, for instance, the development of receptive fields (Clopath et al. [2010]; Song et al. [2000]) or memory and associations is long-standing knowledge. However, the variety of different long-term plasticity rules (Markram et al. [2011]), indicates that the precise synaptic prescriptions of long-term plasticity mechanisms remain unclear. On the contrary, short-term plasticity (STP) is well described in terms of models (Costa et al. [2013]; Hennig [2013]; Le Be' and Markram [2006]; Markram et al. [1998b]; Rinaldi et al. [2008]; Romani et al. [2013]; Rotman and Klyachko [2013]; Testa-Silva et al. [2012]; Tsodyks and Markram [1997]; Varela et al. [1997]) and its role in neuronal computation seems to be related to temporal processing (Carvalho and Buonomano [2011]; Natschläger et al. [2001].

2.3.1 Long-term plasticity

This type of plasticity induces changes in synapses that can last for hours (Gerstner and Kistler [2002b]), which is one of the reasons why it is widely believed that long-term plasticity is at the basis of learning and memory formation (Bliss and Collingridge [1993]; Del Giudice et al. [2003]; Tetzlaff et al. [2012]). Another reason is that this type of plasticity is associated with NMDA receptors that are responsible for calcium influx into the postsynaptic neuron. As mentioned above, Ca^{++} ions are thought to play an important role in long-term

modifications of synaptic strengths (Dayan and Abbott [2001]). Involvement of NMDA receptors, that requires coincidence between presynaptic and postsynaptic activity to be activated, suggests that long-term plasticity is not only a presynaptic or postsynaptic process.

Hebb's original proposal only refers to modifications of synapses leading to strengthening, but already few years after his postulate the idea that connections can also experience depression was starting becoming popular. Similarly to Hebb's formulation, the German biologist *Gunter Stent* (*1924-2008*) proposed in 1973 that depression takes place when two neurons are not sufficiently coactive: *When the presynaptic axon of cell A repeatedly and persistently fails to excite the postsynaptic cell B while cell B is firing under the influence of other presynaptic axons, metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is decreased* (Stent [1973]). In the same decade, long-term depression (LTD) was finally observed in the hippocampus for both inactive pathways and activated pathways provided that the activation frequency is low (Dunwiddie and Lynch [1978]; Lynch et al. [1977]).

The basic original Hebb's formulation formulation has been lately turned into a mathematical relation, called Hebbian rule (Gerstner and Kistler [2002a]):

$$\frac{d}{dt}w_{ij} = \gamma \nu_i \nu_j \tag{2.5}$$

where the constant γ is called learning rate, w_{ij} is the strength of the connection from neuron j to neuron i and ν_i, ν_j are the firing rates.

From this basic rule, many others have been developed, accounting for different features, not only synaptic depression but also for example competition and saturation (Bienenstock et al. [1982]; Gerstner and Kistler [2002a]). Fundamental properties of such a kind of rules are locality and cooperativity: The strength of the connection between i and j depends only on i and j, and not for example on some global signal, and to have potentiation both neurons need to be active.

Strictly related to learning is the coding problem. Since all spikes of a given neuron look alike, it is generally believed that the form of the action potential does not carry any information. The problem of how information is coded and transmitted from neuron to neuron is one of the most challenging in neuroscience. Two main mechanisms have been identified so far, with evidence for both: On one side, information can be transferred through a *rate code*, which is based on the number of spikes per time unit; the second mechanism, instead is based on the exact timing of the spike (Gerstner and Kistler [2002b]), and it is called *temporal* or *spike code*. The Hebbian rule as stated originally by Hebb and as formualated above in Eq. (2.5) is based on a rate code, as it uses firing rates. Hebbian learning can be

written also in terms of the spike code (Abbott and Gerstner [2004]; Gerstner and Kistler [2002a]).

Spike-Timing Dependent Plasticity The most well-know learning mechanism based on the timing of the spikes is called *Spike-Timing Dependent Plasticity (STDP)*. STDP was postulated by the German physicist and neuroscientist *Wulfram Gerstner (1963 -)* in 1996 (Gerstner et al. [1996]) and the first evidence supporting it was found already the year after (Markram et al. [1997]). Experiments showed that to obtain both potentiation and depression of synaptic strength two different protocols were needed: Long-term potentiation occurs whenever the postsynaptic neuron fires after the presynaptic one, vice versa for the long-term depression. These modifications are long-term because they can last even for minutes or more and then it is widely believed that this mechanism is involved in learning and memory formation (Billings and van Rossum [2009]; Song and Abbott [2001]; Vogels et al. [2011]).

Few years later, also the time dependence of this phenomenon has been accurately measured by (Bi and Poo [1998]). Given a presynaptic neuron j and a postsynaptic neuron i, then the STDP rule can be written as:

$$\Delta w_{ij} = \begin{cases} A_+ \exp(-(t_{post} - t_{pre})/\tau_+) & \text{if} \quad (t_{post} - t_{pre}) > 0 \\ A_- \exp(t_{post} - t_{pre}/\tau_-) & \text{if} \quad (t_{post} - t_{pre}) < 0 \end{cases} \tag{2.6}$$

where $A_+ > 0$, $A_- < 0$ and τ_+, τ_- describes the exponential time decay. This rule gives rise to a so called learning window, showed in Fig. (2.2)A with *solid line*. Note that STDP belongs to the family of Hebbian rules, because it satisfies both locality and cooperativity.

The original STDP learning as described by Eq. (2.6) is based on a *pair protocol*, i.e. the learning rule involves only pairs of spikes - one pre and one post. It has been show that it is possible to extend this mechanism to *triplets* of spikes, *pre-post-pre* or *post-pre-post*, and in general to *multiplets* of spikes (Pfister and Gerstner [2006]). It is remarkable that STDP equipped with the triplet rule can capture both rate and temporal codes by changing the firing frequencies of the neurons. At low firing rates the pair-based asymmetric window can be reproduced, obtaining potentiation whenever the pre-post-pre sequence of spikes occurs and depression with the post-pre-post sequence (Fig. 2.2B *Dashed line*). However, as the firing frequency becomes high, we obtain only the pure Hebbian potentiation (Fig. 2.2B *Solid line*).

Despite being very popular and confirmed by experiments, STDP suffers of a stability problem: Without any constraint, strengths of synapses can grow to infinity, which is both not biologically plausible and not optimal from a computational point of view. To avoid it, different techniques have been developed, some more biologically realistic and some others less (Babadi and Abbott [2010]; van Rossum et al. [2000]). The most simple solution is to

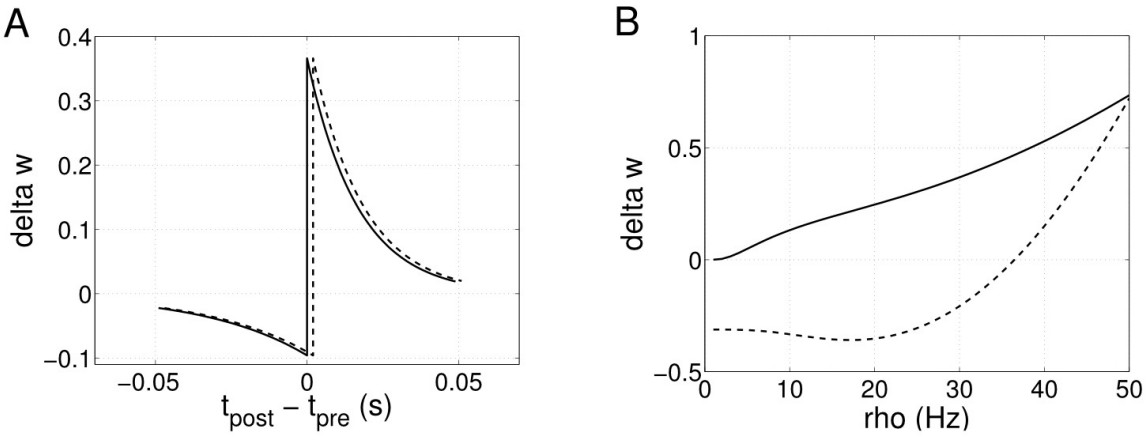

Figure 2.2 Spike-Timing Dependent Plasticity learning rules. **A** Learning window with a pair rule. *Solid line*: Reproduction of standard pair-based rule. *Dashed line*: Reproduction of the shifted rule (Babadi and Abbott [2010]) with $d = 2ms$. **B** Frequency effect with the triplet rule from Pfister and Gerstner [2006]. *Solid line*: Post-pre-post firing protocol, giving potentiation. *Dashed line*: Pre-post-pre firing protocol, giving depression at low rates (pure STDP) and potentiation at high rates (pure Hebbian) [figure produced with MATLAB®].

define an upper and lower bound for the weights (Gerstner and Kistler [2002b]). Because of the competition hidden in the rule, at the end of the evolution some of the connections collapse to the maximum value and the others to the minimum value, giving the characteristic bimodal distribution (Song et al. [2000]).

A more interesting technique consists in shifting the learning window in such a way when the postsynaptic neuron fires a spike right after the presynaptic neuron there is depression rather than potentiation (Fig. 2.2A *dashed line*). The reason of why the shift stabilizes the evolution is very well explained and also this procedure is well justified from a biological point of view, being related to the slow activation of NMDA receptors on the postsynaptic neuron (Babadi and Abbott [2010]).

2.3.2 Short-term plasticity

Based solely on STDP, when a postsynaptic neuron is repeatedly stimulated with a train of presynaptic spikes, an increase in the amplitude of the EPSP should appear. However, this is not always the case, as another mechanism is indeed present at synaptic level, called short-term plasticity as it acts on time scales shorter than the ones involved in long-term plasticity. Suggested biological mechanisms underlying STP are changes in the vesicles release probability and accumulation of calcium in the synaptic tterminal. Both of them

affect the amount of neurotransmitters released by the presynaptic neuron, suggesting that STP is a pure presynaptic mechanism, contrarily to long-term plasticity.

Depression it has been observed in neocortical pyramidal neurons that synaptic responses to presynaptic stimulation strongly depend on the synaptic activity, i.e. on the history of the synapse (Gupta et al. [2000]; Markram et al. [1998b]). In particular, connections between pyramidal neurons typically display pronounced *synaptic depression*, characterized by fast decrease of synaptic response during the presynaptic stimulation (Markram et al. [1998b]). This behaviour can be associated with the availability of resources when the presynaptic signal arrives at the synapse: to evoke an EPSP a neuron has to use some resources which afterwards become temporarily unavailable to the cell; if another presynaptic spike arrives before these resources are fully restored, the response of the neuron will be less efficient, displaying fatigue.

This formulation can be translated into a system of kinetic equations. According with one well-established phenomenological model (Tsodyks and Markram [1997]), resources can be found in three states: *Recovered*, *active* and *inactive*. Let x, y, z be the fractions of resources in each of these states. If all the resources are activated by a presynaptic action potential, then the maximal possible response would be observed, defined as the absolute synaptic efficacy A. In reality, each presynaptic action potential utilizes only a fraction u of resources currently available in the recovered state, which therefore instantly move from recovered to active state. Following this, these resoures quickly inactivate with a time constant τ_{in} of few milliseconds and finally recover with a larger time constant τ_{rec} of $\sim 1\,s$. If we denote with t_{sp} the time of spike arrival then we can write the dynamical equations (Chow et al. [2005]) as:

$$\begin{cases} \frac{dx}{dt} &= \frac{z}{\tau_{rec}} - ux\delta\left(t - t_{sp}\right) \\ \frac{dy}{dt} &= -\frac{y}{\tau_{in}} + ux\delta\left(t - t_{sp}\right) \\ \frac{dz}{dt} &= \frac{y}{\tau_{in}} - \frac{z}{\tau_{rec}} \end{cases} \tag{2.7}$$

Every time the presynaptic spike arrives, it triggers a postsynaptic current *PSC* with amplitude $E = Aux$, proportional to the amount of resources in the recovered state immediately before the spike arrival time, and with time decay τ_{in}. At every time, the postsynaptic current is proportional to the fraction of resources in the active state, $Ay(t)$. Note that the higher the u the faster synaptic resources are utilized, which leads to more rapid depression. This formulation ignores the stochastic nature of synaptic release and reproduces the average postsynaptic responses generated by any presynaptic spike train for inter-pyramidal synapses (Tsodyks and Markram [1997]).

The above set of equations can often be reduced to a simpler one, due to a very short time constant for the inactive state. Indeed, when $\tau_{in} \geq \tau_{rec}$, the variable z can be neglected, meaning that $x = 1 - y$. The 3-variable model can therefore be reduced to a single-variable model (Chow et al. [2005]):

$$\frac{dx}{dt} = \frac{1-x}{\tau_{rec}} - ux\delta\left(t - t_{sp}\right) \tag{2.8}$$

Facilitation Short-term synaptic depression is not the only form of activity dependence of synaptic response. Connections in neocortex involving pyramidal neurons and interneurons – small, locally projecting (in contrast to larger projection neurons with long-distance connections) and typically inhibitory – usually exhibit various degree of *synaptic facilitation*, characterized by fast increase of synaptic response during the pre-synaptic stimulation (Gupta et al. [2000]). In this case, the interpretation is given in term of time required by the synapse to "charge" itself to its full capacity: It responds weakly to the first spike and then progressively stronger to the following ones, depending on the recovery time of the resources.

Eq. (2.8) describes only short-term synaptic depression. Facilitation can be easily taken into account by assuming that the value of u is no longer fixed, but rather increases with each presynaptic spike and returns to baseline with a time constant τ_{facil}. Such an increase is determined by a parameter U (Chow et al. [2005]; Morrison et al. [2008]):

$$\frac{du\_}{dt} = -\frac{u\_}{\tau_{facil}} + U\left(1 - u\_\right)\delta\left(t - t_{sp}\right) \tag{2.9}$$

In the above equation, the notation $u\_$ means that the value of u has to be taken right before the arrival of the spike, as in $u(t-\varepsilon)$ with ε very small. The value of u used in Eq. (2.8) (as well as in Eq. (2.7)) is instead the value of u right after the spike. These two quantities are related by the update rule following a spike arrival:

$$u = u\_\left(1 - U\right) + U \tag{2.10}$$

In this description, u represents the amount of resources effectively used by a spike: when the first spike arrives at the presynaptic terminal, $u\_ = 0$ (stationary state of Eq. (2.9)) and all the resources are available to the cell (stationary state of Eq. (2.8)); hence $u = U$ is the fraction of resources that will be used in the transmission of the signal (Eq. (2.10)), which will cause a decrease in the fraction of resources in the recovered state from $x = 1$ to $x = 1 - U$ (Eq. (2.8)). Following this, in the absence of other spikes, x increases back to 1 whereas $u\_$ decreases to zero and consequently u decreases to U. From this description it

emerges how $u\_$ can be interpreted as the amount of used resources, whereas u represents the amount of resources that are going to be used if a spike arrives in that moment; in other words, u represents the same quantity as $u\_$ but before the effective usage of resources. To avoid this unnecessary double variable notation within the model, Eq. (2.10) can be substituted into Eq. (2.9), resulting in the following:

$$\frac{du}{dt} = -\frac{u-U}{\tau_{facil}} + U\,(1-u)\,\delta\,(t-t_{sp}) \tag{2.11}$$

Interplay Facilitation and depression are strictly interconnected since stronger facilitation leads to higher values of u, which in turn leads to stronger depression. The solutions of the above equations (Chow et al. [2005]) show that, when the presynaptic neuron emits regular spike trains with a frequency f, the EPSP reaches a frequancy-dependent steady-state, both when the synapse is only depressing and when there is also facilitation: when the synapse shows only depression then the stationary amplitude of an EPSP is in inverse proportion with f, whereas when facilitation is added then the plateau amplitude exhibit a non-monotonic dependency. This is intuitive: When facilitation is too high, the amount of resources consumed grows and if the frequency of stimulation is above a certain threshold the cell does not have the time to restore resources, hence synaptic depression is activated. Therefore, what determines the response of the synapse is the value of U: When it is small then facilitation is marked, whereas for higher values it is not observed. This model well captures the main features of synaptic transmission between pyramidal neurons and interneurons when $U \sim 0.01 \rightarrow 0.05$.

Biological interpretation According with the above derivation, the variables x and u correspond to precise biological quantities, respectively: occupancy of the release pool and probability of neurotransmitter release, both related to the presynaptic cell (Hennig [2013]).

The presynaptic terminal contains a specialised area where vesicles filled with the neurotransmitters required for the signal transmission are clustered. These vesicles are organised into anatomically distinguishable populations, called vesiscles pools: the release pool, containing vesicles ready to release, and the reserve pool, where the remaining vesicles are on hold and ready to repleace empty vesicles in the release pool following a neurotransmitter rerelease event. Calcium ions are the agents causing vesicles to migrate from the release pool towards the presynaptic membrane, where they merge with it releasing the neurotransmitters into the synaptic cleft. The release of a single vecicle constitutes the smallest signal, or the quantum, that can be transmitted to the postsynaptic cell. This single process is characterised by a release probability $p(t)$. If $N(t)$ is the number of vesicles available for release (i.e.

present in the release pool), then the amount of neurotransmitters released into the synaptic cleft is:

$$T(t) = p(t)N(t) \quad . \tag{2.12}$$

The arrival of a presynaptic action potential at the axon terminal triggers the opening of voltage gated calcium channels with subsequent influx of extracellular Ca^{++} by diffusion. This leads to an increase in the number of vesicles driven towards the cell membrane, and therefore in the number of released neurotransmitters.

According with this description, the time constant for synaptic depression τ_{rec} describes how fast the release pool is replenished with vesicles from the reserve pool. Becuase the release probability is ultimately regulated by excess in calcium concentration, the time constant for facilitaiton τ_{rec} can be associated with the speed at which calcium ions concentration decays back to its resting level. Finally, the interpretation of U is rather straightforward: as it describes the fraction of resources used by the firts action potential, it corresponds to the initial (resting) value of release probability, before that incresed calcium concentration due precisely to that spike brings it to higher values.

The amount of neurotransmitters realeased $T(t)$ is responsible for the transmisison of the signal from the presynaptic terminal to the postsynaptic neuron, where they binds to receptors to generate the postsynaptic response. Specifically, an increase in the rate of binding events determines an increase in the conductance g of the postsynaptic neuron, which corresponds to a depolarisation. Therefore the simplest way of modelling of this process is with the following linear relationship:

$$g(t) \propto T(t) \quad , \tag{2.13}$$

which clearly shows that the postsynaptic responce is related to the amount of neurotransmitters release by the presynaptic neuron. Therefore, facilitation and depression acts on $T(t)$. The most accepted mechanism for facilitation involves accumulation of calcium ions after a presynaptic spike, with subsequent increase in the release probability $p(t)$. Synaptic depression, instead, is due to a depletion of vesicles in the release pool, hence to a decrease of the number of available vescicles $N(t)$ (or, equivalently, of the occupancy of the release pool) (Hennig [2013]).

Summary In this Chapter some basic elements from the literature were given, such as how the activity of a neuron is modelled and, most important, how it affects the transmission itself and the connections between neurons. Both mechanisms of synaptic modification were introduce, which will be largely used in the rest of the thesis. In the next three chapters the three papers I produced will be presented, starting with the two papers addressing the question how we can detect motifs.

Chapter 3

Measuring Symmetry, Asymmetry and Randomness in Neural Network Connectivity[1]

Studies on the brain wiring diagram have shown that connectivity is non-random, highlighting the existence of specific connectivity motifs at the microcircuit level, see for instance Perin et al. [2011]; Silberberg and Markram [2007]; Song et al. [2005]; Wang et al. [2006]. Of particular interest are the motifs that exhibit bidirectional (reciprocal) and unidirectional (non-reciprocal) connections between pairs of neurons. More specifically, theoretical work (Clopath et al. [2010]) studied the development of unidirectional connectivity due to long-term plasticity in an artificial network of spiking neurons under a *temporal coding scheme*, where it is assumed that the time at which neurons fire carries out important information. This finding is correlated to unidirectional connectivity observed in somatosensory cortex, see Lefort et al. [2009]. In Clopath et al. [2010] the development of bidirectional connectivity in the same network under a *frequency coding scheme*, where information is transmitted in the firing rate of the neurons, was also studied and correlated to bidirectional connectivity found in the visual cortex (Song et al. [2005]). Complementary to this work, in Vasilaki and Giugliano [2012, 2014] the authors explored the experimentally identified correlation of bidirectional and unidirectional connectivity to short-term synaptic dynamics, see Pignatelli [2009], by studying the development of connectivity in networks with facilitating and depressing synapses due to the interaction of short-term and long-term plasticities. The role of synaptic long-term plasticity in structures formation within networks has been also investigated in Babadi and Abbott [2013]; Bourjaily and Miller [2011a,b].

[1] This chapter is almost entirely taken from Esposito et al. [2014]

Similar to Clopath et al. [2010] and Vasilaki and Giugliano [2012, 2014], we hypothesise that the above mentioned motifs have been shaped via activity dependent synaptic plasticity processes, and that learning moves the distribution of the synaptic connections away from randomness. Our aim is to provide a global, macroscopic, single parameter characterisation of the statistical occurrence of bidirectional and unidirectional motifs. To this end:

1. We define a symmetry measure that does not require any a priori thresholding of the weights or knowledge of their maximal value, and hence is applicable to both simulations and experimental data.

2. We calculate the mean and variance of this symmetry measure for random uniform or Gaussian distributions, which allows us to introduce a confidence measure of how significantly symmetric or asymmetric is a specific configuration, i.e. how likely it is that the configuration is the result of chance.

3. We demonstrate the discriminatory power of our symmetry measure by inspecting the eigenvalues of different types of connectivity matrices, given that symmetric matrices are known to have real eigenvalues.

4. We show that a Gaussian distribution biases the connectivity motifs to more symmetric configurations than a uniform distribution and that introducing a random synaptic pruning, mimicking developmental regulation in synaptogenesis, biases the connectivity motifs to more asymmetric configurations, regardless of the distribution. Our statistics of the symmetry measure allows us to correctly evaluate the significance of a symmetric or asymmetric network configuration in both these cases.

5. Our symmetry measure allows us to observe the evolution of a specific network configuration, as we exemplify in our results.

We expect that our work will benefit the computational modelling community, by providing a systematic way to characterise symmetry and asymmetry in network structures. Further, our symmetry measure will be of use to electrophysiologists that may investigate symmetric or asymmetric network connectivity.

3.1 Methods

In what follows, we first define a novel measure that quantifies the degree of symmetry in a neuronal network with excitatory synaptic connections. More specifically, we describe the strength of the synaptic efficacies between the neurons by the elements of a square matrix,

i.e. the connectivity matrix, to which we associate a number that quantifies the similarity of the elements above the matrix diagonal to those below the diagonal. We further study this measure from a statistical point of view, by means of both analytical tools and numerical simulations. Aiming to associate a significance value to the measure, i.e. the probability that a certain symmetric or non-symmetric configuration is the result of chance, we consider random synaptic efficacies drawn from uniform and Gaussian distributions. We also study how our symmetry measure is affected by the anatomical disconnection of neurons in a random manner, i.e. zeroing some entries in the connectivity matrix. Finally, we anticipate that connectivity distributions are modified by activity-dependent processes and we describe the structure of the network we use as a demonstrative example in the Results section.

3.1.1 Definitions

Let us consider the adjacency (or connectivity) matrix W of a weighted directed network (Newman [2010]), composed of N vertices and without self-edges. The N vertices represent the neurons, with $N(N-1)$ possible synaptic connections among them. The synaptic efficacy between two neurons is expressed as a positive element w_{ji} in the adjacency matrix. W is thus composed by positive elements off-diagonal, taking values in the bounded range $[0, w_{max}]$, and by zero diagonal entries. We define s as a measure of the symmetry of W:

$$s = 1 - \frac{2}{N(N-1) - 2M} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{\left| w_{ij} - w_{ji} \right|}{w_{ij} + w_{ji}} \tag{3.1}$$

where M is the number of instances where both w_{ij} and w_{ji} are zero, i.e. there is no connection between two neurons. The term $q = \frac{N(N-1)-2M}{2}$ is a normalisation factor that represents the total number of synaptic connection pairs that have at least one non-zero connection. A value of s near 0 indicates that there are virtually no reciprocal connections in the network, while a value of s near 1 indicates that virtually all connections are reciprocal. We exclude (0,0) pairs from our definition of the symmetry measure. Mathematically such pairs would introduce undefined terms to Eq. (3.1). In addition, conceptually, we expect that small weights will not be experimentally measurable. It is then reasonable to exclude them, expecting to effectively increase the signal to noise ratio.

Pruning and Plasticity. We assume that a connection w_{ij} is permanently disconnected and set to 0 with probability $\wp_{pruning} = a \in [0, 1)$. Consequently, the probability that two neurons i and j are mutually disconnected, i.e. $w_{ij} = w_{ji} = 0$, is a^2. When a connection is permanently *pruned* in such a way, its efficacy remains 0 all the time, whereas the off-diagonal non-pruned values of the adjacency matrix W change slowly in time, as a result

of activity-dependent synaptic plasticity. We consider that this procedure correlates with developmental mechanisms associated with or following synaptogenesis.

Unidirectional and Bidirectional connection pairs. We associate the quantity $Z_{ij} = \frac{|w_{ij}-w_{ji}|}{w_{ij}+w_{ji}}$, i.e. the term of the summation in the Eq. (3.1) to the neuronal pair i,j. This term maps the strength of the connections between two neurons to a single variable. Each connection pair can therefore be bidirectional if $w_{ij} \simeq w_{ji}$, unidirectional if $w_{ij} \ll w_{ji}$ or $w_{ij} \gg w_{ji}$, or none of the two. As a consequence, a network can be dominated by bidirectional connectivity, by unidirectional connectivity, or it may exhibit random features.

Weight Bounds. In what follows we consider the case of $w_{max} = 1$. Due to the term $\frac{|w_{ij}-w_{ji}|}{w_{ij}+w_{ji}}$, this can be done without loss of generality.

Inhibitory Networks. The symmetry measure defined in Eq. (3.1) is based on the assumption that synaptic weights are positive, hence that the network is formed only by excitatory neurons. It is easy to extend the valdity of the measure to networks with only inhibitory neurons, for which $w_{ij} \in [-1,0]\ \forall i,j$:

$$s_{inh} = 1 - \frac{2}{N(N-1)-2M} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{|w_{ij} - w_{ji}|}{|w_{ij}| + |w_{ji}|}. \tag{3.2}$$

In the case of an excitatory network, Eq. (3.2) reduces to Eq. (3.1). In the case of an inhibitory network, everything works symmetrically: a pair of neurons having reciprocal connections of significative different strength will still form a unidirectional pair, whereas if their connections have a similar weight they will still form a bidirectional pair. Therefore, all the following results holds true for inhibitory networks as well, provided a mirroring around zero in the weights domain.

3.1.2 Statistics of s

Let us consider a large number of n instances of a network whose connection weights are randomly distributed. Each adjacency matrix can be evaluated via our symmetry measure. We rewrite Eq. (3.1) as:

$$s = 1 - \frac{1}{q} \sum_{i=1}^{N} \sum_{j=i+1}^{N} Z_{ij} = 1 - \frac{1}{q} \sum_{k=1}^{q} Z_k, \tag{3.3}$$

where k is a linear index running over all the q non-zero "connection pairs" within the network. We can then estimate the mean μ_s and variance σ_s^2 of s over all n networks as:

$$\mu_s \equiv \mathbb{E}_n[s] = 1 - \mathbb{E}_n[Z] \tag{3.4}$$

$$\sigma_s^2 \equiv \mathrm{Var}_n(s) = \frac{1}{q}\mathrm{Var}_n(Z) \tag{3.5}$$

where the notation $\mathbb{E}_n[\cdot]$ and $\mathrm{Var}_n(\cdot)$ implies that the expected value and variance are computed along the n different representations of the network.

Eq. (3.4), (3.5) allow us to transfer the statistical analysis from s to Z. To derive theoretical formulas for mean value and variance of Z we use the fact that its probability density function (PDF), $f(Z)$, can be written as a joint distribution, $f(Z_1, Z_2)$ where we have introduced the notation $Z_1 = |w_{ij} - w_{ji}|$, $Z_2 = w_{ji} + w_{ji}$:

$$\mathbb{E}[Z] = \int_0^1 dZ\, Z f(Z) = \int\int_D dZ_1\, dZ_2\, \frac{Z_1}{Z_2} f(Z_1, Z_2) \tag{3.6}$$

within the range D defined by $0 \leq Z_1 \leq 1$ and $Z_1 \leq Z_2 \leq 2 - Z_1$. Similarly, we can calculate the variance as follows:

$$\mathrm{Var}(Z) = \int_0^1 dZ\, (Z - \mathbb{E}[Z])^2 f(Z) = \int\int_D dZ_1\, dZ_2 \left(\frac{Z_1}{Z_2} - \mathbb{E}[Z]\right)^2 f(Z_1, Z_2)\,. \tag{3.7}$$

We note that mean value and variance of Z can be numerically estimated either by using a large set of small networks or on a single very large network: What matters is that the total number of connection pairs, given by the product $n \cdot q$, is sufficiently large to guarantee good statistics and that connection pairs are independent of each other. In the calculations below, we assume a very large adjacency matrix.

3.1.3 Adjacency matrix with uniform random values

We first consider a network with randomly distributed connections without pruning, followed by the more general case where pruning is taken into account.

Fully connected network For the uniform distribution $f^u(w) = 1$ for $w \in [0, 1]$, see Fig. 3.1A. The probability of having $w_{ij} = w_{ji} = 0$ for at least one pair (i, j) is negligible, hence $M = 0$. It is straightforward to derive the distributions $f_2^u(Z_2)$ and $f_1^u(Z_1)$, depicted in Fig. 3.1B,C correspondingly:

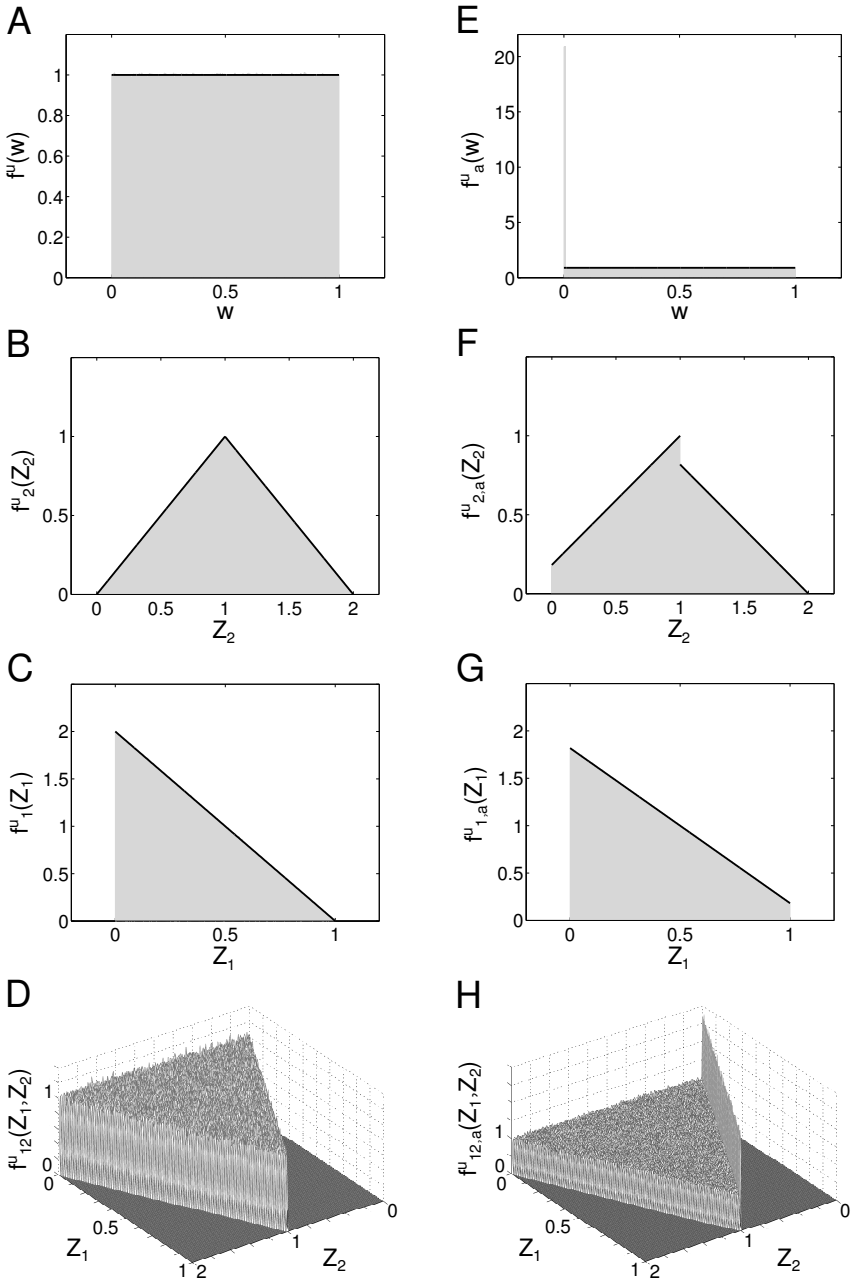

Figure 3.1 Probability density functions for the case of uniformly distributed connections. **A** Distribution of the uniform variable w. **B** Distribution of the sum Z_2 of two uniform variables. **C** Distribution of the absolute difference Z_1 of two uniform variables. **D** Joint distribution of Z_2 and Z_1. **E - G** The same as A, B and C but with pruning $a = 0.1$. **H** The same as D but with pruning $a = 0.01$. In all figures, *Grey shaded area*: Histograms from simulations, *Black lines and surfaces*: Theoretical results (see Eq. (3.8)-(3.14)).

$$f_1^u(Z_1) = \begin{cases} -2Z_1 + 2 & \text{for} \quad Z_1 \in [0,1] \\ 0 & \text{otherwise} \end{cases} \tag{3.8}$$

$$f_2^u(Z_2) = \begin{cases} Z_2 & \text{for} \quad Z_2 \in [0,1] \\ -Z_2 + 2 & \text{for} \quad Z_2 \in [1,2] \\ 0 & \text{otherwise} \end{cases} . \tag{3.9}$$

We can therefore obtain the joint PDF (Fig. 3.1D):

$$f_{12}^u(Z_1, Z_2) = \begin{cases} 1 & \text{for} \quad [Z_1, Z_2] \in D \\ 0 & \text{otherwise} \end{cases} . \tag{3.10}$$

Pruning Introducing pruning to the elements of the adjacency matrix, with probability a, corresponds to a discontinuous probability distribution function of w, that can be written as a sum of a continuous function and of a Dirac's Delta centred in $w = 0$ (see also Fig. 3.1E):

$$f_a^u(w) = (1-a) f^u(w) + a\,\delta(0) . \tag{3.11}$$

Now the $(0,0)$ pairs have to be explicitly excluded from the distributions of Z_1 and Z_2. Also, the number of pairs of the type $(w,0)$ increases, resulting in the appearance of a uniform contribution in the region $[0,1]$ in both the PDF of Z_1 and Z_2. Their final exact profile can be obtained by considering the possible combinations of drawing w_{ij} and w_{ji} from the above pruned distribution and their corresponding probability of occurrence. There are four contributions: $f^u(w) \times f^u(w)$, $f^u(w) \times \delta(0)$, $\delta(0) \times f^u(w)$, $\delta(0) \times \delta(0)$. The last term, which describes the $(0,0)$ pairs, has to be subtracted and the remaining expression has to be renormalised. The results are graphically shown in Fig. 3.1F, 3.1G and are mathematically described by the following expressions:

$$f_{1,a}^u(Z_1) = \begin{cases} -2\dfrac{1-a}{1+a} Z_1 + \dfrac{2}{1+a} & \text{for} \quad Z_1 \in [0,1] \\ \\ 0 & \text{otherwise} \end{cases} \tag{3.12}$$

$$f^u_{2,a}(Z_2) = \begin{cases} \dfrac{1-a}{1+a} Z_2 + \dfrac{2a}{1+a} & \text{for} \quad Z_2 \in [0,1] \\ -\dfrac{1-a}{1+a} Z_2 + 2\dfrac{1-a}{1+a} & \text{for} \quad Z_2 \in [1,2] \\ 0 & \text{otherwise} \end{cases} \tag{3.13}$$

The joint PDF is a mixture of two uniform distributions: The unpruned distribution $f^u_{12}(Z_1, Z_2)$ and the contribution from the pruning, $f^u_{peak}(Z_1, Z_2)$, which is a delta peak along the line $Z_1 = Z_2$, see Fig. 3.1H. To obtain $f^u_a(w)$, the two unitary distributions are mixed with some coefficients c_2 and c_1, satisfying the normalisation condition $c_1 + c_2 = 1$. With the same arguments used for $f^u_{1,a}(Z_1)$ and $f^u_{2,a}(Z_2)$, we can derive the relation between c_1, c_2 and a, so that we can finally write:

$$f^u_{12,a}(Z_1, Z_2) = c_1\, f^u_{12}(Z_1, Z_2) + c_2\, f^u_{peak}(Z_1, Z_2) = \frac{1-a}{1+a} + \frac{2a}{1+a}\,\delta(Z_1 - Z_2) \quad \text{for } [Z_1, Z_2] \in D\,. \tag{3.14}$$

Expected value and variance of Z

We can calculate mean value and variance of Z by plugging Eq. (3.14) into Eq. (3.6) and (3.7):

$$\mathbb{E}^u[Z] = \frac{1-a}{1+a}\Big(2\ln 2 - 1\Big) + \frac{2a}{1+a} \tag{3.15}$$

$$\mathrm{Var}^u(Z) = 1 - \frac{8\ln 2\,(1-a) + 7a - 5}{1+a} - \left(\frac{2\,(1-a)\,(1-\ln 2)}{1+a}\right)^2. \tag{3.16}$$

Expected value and variance of s

By combining the above results with Eq. (3.4) and (3.5), we can derive the final formulas for the expected value and variance of s:

$$\mu^u_s \equiv \mathbb{E}^u[s] = 1 - \frac{1-a}{1+a}\Big(2\ln 2 - 1\Big) - \frac{2a}{1+a} \tag{3.17}$$

$$\left(\sigma^2\right)_s^u \equiv \mathrm{Var}^u(s) = \frac{1}{q}\left[1 - \frac{8\ln 2\,(1-a)+7a-5}{1+a} - \left(\frac{2\,(1-a)\,(1-\ln 2)}{1+a}\right)^2\right] \tag{3.18}$$

3.1.4 Adjacency matrix with Gaussian-distributed random values

The procedure described above to derive the joint PDF of Z_1 and Z_2 is applicable to any distribution. In what follows, we consider a network with initial connections drawn by a truncated Gaussian distribution.

Distribution of connections. Whereas the uniform distribution is well defined in any finite interval, the Gaussian distribution requires some considerations. Strictly speaking, any Gaussian distribution is defined over the entire real axes. For practical reasons, however, for any finite network $N < \infty$, the maximum and the minimum values of the weights, w_{max} and w_{min}, are always well defined, and therefore the actual distribution is a truncated Gaussian. To be able to consider the truncated Gaussian distribution as Gaussian with satisfactory accuracy, we require that the portion of the Gaussian enclosed in the region $[w_{min}, w_{max}]$ is as close as possible to 1. This means that the distribution has to be narrow enough with respect to the interval of definition $[w_{min}, w_{max}]$. Also, by definition, the distribution has to be symmetric in $[w_{min}, w_{max}]$. Because we are considering only excitatory connections then $w_{min} = 0$, so as the mean value has to be $\mu_w = \frac{w_{max}}{2}$. Additionally, the narrowness imposes a condition on the standard deviation of the distribution: $\sigma_w \ll \Delta w = w_{max}$. Since we can set $w_{max} = 1$ without loss of generalization, the entire study on all the possible Gaussian distributions can be limited to a special class, $\mathscr{G}\left(\mu = \frac{1}{2}, \sigma \ll 1\right)$.

The choice of σ. To guarantee a good approximation of a Gaussian distribution, we define the truncated Gaussian distribution such that points within 5σ fall in [0,1] leading to $\sigma_w = \frac{1}{10}$ and a truncation error $\varepsilon_{tr} = 1 - \int_0^1 \mathscr{N}(w)\,dw \simeq 10^{-4}$.

Fully connected network For the truncated Gaussian distribution defined above, the distribution of connections without pruning is (see also Fig. 3.2A):

$$f^g(w) = \mathscr{N}\left(w; \mu_w = \frac{1}{2}, \sigma_w = \frac{1}{10}\right) = \frac{1}{\sqrt{2\pi\,0.1^2}} \exp\left(-\frac{(w-0.5)^2}{2\times 0.1^2}\right) \quad \text{for } w \in [0,\,1] \tag{3.19}$$

where $\mathscr{N}$ denotes the normal distribution. Since combinations of Gaussian distributions are also Gaussian distributions, we can immediately derive the PDF of Z_2 and $Z_1^* = w_{ij} - w_{ji}$. Then, $f_1^g(Z_1)$ is simply the positive half of $f_1^{*g}(Z_1^*)$, but scaled by a factor of two because of the normalization. We obtain (Fig. 3.2B,C):

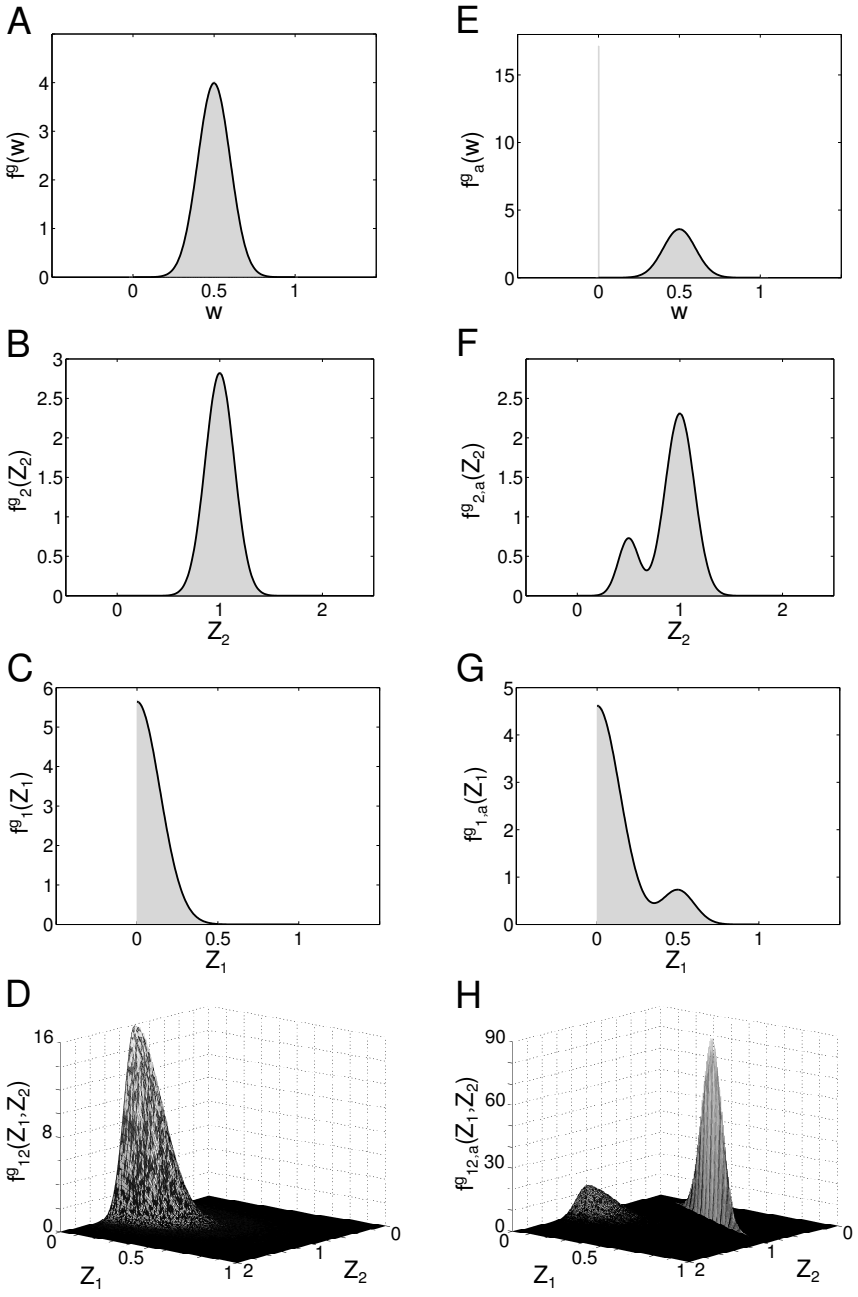

Figure 3.2 Probability density functions for the case of Gaussian-distributed connections. **A** Distribution of the Gaussian variable w. **B** Distribution of the sum Z_2 of two Gaussian-distributed variables. **C** Distribution of the absolute difference Z_1 of two Gaussian-distributed variables. **D** Joint distribution of Z_2 and Z_1. **E - H** The same as A - D but with pruning $a = 0.1$. In all the figures, *Grey shaded area*: Histograms from simulations, *Black lines and surfaces*: Theoretical results (see Eq. (3.20)-(3.26)).

$$f_1^g(Z_1) = \mathscr{N}_{1/2}\left(Z_1; \mu_1 = 0, \sigma_1 = \sqrt{2}\sigma_w\right) \quad \text{for } Z_1 \in [0, 1] \tag{3.20}$$

$$f_2^g(Z_2) = \mathscr{N}\left(Z_2; \mu_2 = 2\mu_w, \sigma_2 = \sqrt{2}\sigma_w\right) \quad \text{for } Z_2 \in [0, 2] \tag{3.21}$$

where $\mathscr{N}_{1/2}$ identifies the normalised (positive) half of a normal distribution. Similarly, the joint distribution $f_{12}^g(Z_1, Z_2)$ can be easily derived from the bivariate Gaussian of Z_1^* and Z_2 (Fig. 3.2D):

$$f_{12}^g(Z_1, Z_2) = \mathscr{N}_{1/2}^B\left(Z_1, Z_2; \boldsymbol{\mu}_{1,2}, \boldsymbol{\Sigma}_{1,2}\right) \tag{3.22}$$

with $\mathscr{N}_{1/2}^B$ being the normalised half (where $Z_1 > 0$) of a bivariate normal distribution.

Pruning When taking pruning into account, each PDF can be considered as a mixture of the unpruned distribution and the contribution coming from the pruning. We can therefore write:

$$f_a^g(w) = (1-a)\,\mathscr{N}(w; \mu_w, \sigma_w) + a\,\delta(0)\,, \tag{3.23}$$

$$f_{1,a}^g(Z_1) = \frac{1-a}{1+a}\,\mathscr{N}_{1/2}(Z_1; \mu_1, \sigma_1) + \frac{2a}{1+a}\,\mathscr{N}(w; \mu_w, \sigma_w) \quad \text{for } Z_1 \in [0, 1] \tag{3.24}$$

$$f_{2,a}^g(Z_2) = \frac{1-a}{1+a}\,\mathscr{N}(Z_2; \mu_2, \sigma_2) + \frac{2a}{1+a}\,\mathscr{N}(w; \mu_w, \sigma_w) \quad \text{for } Z_2 \in [0, 2] \tag{3.25}$$

The above distributions are plotted in Fig. 3.2E, 3.2F, 3.2G.

Finally, the joint PDF is again a mixture model, with a univariate Gaussian peak profile on the line $Z_1 = Z_2$ (Fig. 3.2H). Note that this peak can be described by the intersection of the plane $Z_1 = Z_2$ with the full unpruned bivariate normal distribution $\mathscr{N}^B$ transformed to have its mean in (μ_w, μ_w). This operation implies a re-normalisation by $\mathscr{Z} = \left(4\pi\sigma_2^2\right)^{-1/2}$ of the resulting univariate Gaussian. Then, we can write:

$$
\begin{aligned}
f^{g}_{12,a}(Z_1, Z_2) = & \frac{1-a}{1+a}\,\mathscr{N}^{B}_{1/2}\left(Z_1, Z_2; \boldsymbol{\mu}_{1,2}, \boldsymbol{\Sigma}_{1,2}\right) \\
& + \frac{2a}{1+a}\,\frac{1}{\mathscr{Z}}\,\mathscr{N}^{B}\left(Z_1, Z_2; \boldsymbol{\mu}_{w,w}, \boldsymbol{\Sigma}_{1,2}\right)\,\delta(Z_1 - Z_2) \quad \text{for } [Z_1, Z_2] \in D
\end{aligned}
\tag{3.26}
$$

Correlation in the bivariate Gaussian The correlation ρ between Z_1 and Z_2, appearing in the off-diagonal terms of $\boldsymbol{\Sigma}_{1,2}$, can be computed by running a numerical simulation. We estimated ρ as a mean value over 10^5 representations of a 10-neuron network with random connections distributed according to $\mathscr{N}(w;\, \mu_w,\, \sigma_w)$ and with no pruning, i.e. $a = 0$. The result is $\rho \simeq 7 \times 10^{-4}$, which allows to treat Z_1 and Z_2 as independent variables and then to factorise the bivariate normal distribution Eq. (3.22) in the product of the two single distributions. Indeed, by introducing the Heaviside step function $\Theta(x)$ and the re-normalisation parameter $R = 2$, we can write:

$$
\begin{aligned}
\mathscr{N}^{B}_{1/2}\left(Z_1, Z_2; \boldsymbol{\mu}_{1,2}, \boldsymbol{\Sigma}_{1,2}\right) &= R\,\mathscr{N}^{B}\left(Z_1, Z_2; \boldsymbol{\mu}_{1,2}, \boldsymbol{\Sigma}_{1,2}\right)\Theta\left(Z_1 - \mu_1\right) \\
&\simeq R\,\mathscr{N}\left(Z_1;\, \mu_1,\, \sigma_1\right)\mathscr{N}\left(Z_2;\, \mu_2,\, \sigma_2\right)\Theta\left(Z_1 - \mu_1\right) \\
&= \mathscr{N}_{1/2}\left(Z_1;\, \mu_1,\, \sigma_1\right)\mathscr{N}\left(Z_2;\, \mu_2,\, \sigma_2\right).
\end{aligned}
\tag{3.27}
$$

We note that the pruning case does not require a different calculation and can be treated as the $a = 0$ case. This is because we are describing the effect of the pruning with a separate (univariate) function, i.e. the halved bivariate normal distribution describes only the unpruned part of the network, see Eq. (3.26).

The suitability of this approximation is also certified by Fig. 3.2D,H, where the agreement between simulation results and theoretical fit with Eq. (3.27) is excellent.

Expected value and variance of Z

Now we can insert the expression of the joint distribution, Eq. (3.26), into Eq. (3.4) and Eq. (3.5):

$$
\mathbb{E}^{g}[Z] = \frac{1-a}{1+a}\int_0^1 dZ_1\, Z_1 \int_{Z_1}^{2-Z_1} dZ_2\, \frac{1}{Z_2}\,\mathscr{N}^{B}_{1/2}\left(Z_1, Z_2; \boldsymbol{\mu}_{1,2}, \boldsymbol{\Sigma}_{1,2}\right) + \frac{2a}{1+a}
\tag{3.28}
$$

$$\begin{aligned}\mathrm{Var}^g(Z) = \frac{1-a}{1+a}\int_0^1 dZ_1 \int_{Z_1}^{2-Z_1} dZ_2 \left(\frac{Z_1}{Z_2} - \mathbb{E}^g[Z]\right)^2 \mathscr{N}^B_{1/2}\left(Z_1, Z_2; \boldsymbol{\mu}_{1,2}, \boldsymbol{\Sigma}_{1,2}\right) \\ + \frac{2a}{1+a}\frac{1}{\mathscr{Z}}\int_0^1 dZ_1 \left(1 - \mathbb{E}^g[Z]\right)^2 \mathscr{N}^B\left(Z_1, Z_1; \boldsymbol{\mu}_{w,w}, \boldsymbol{\Sigma}_{1,1}\right)\end{aligned} \tag{3.29}$$

To calculate the above expression we use symbolic integration.

Expected value and variance of s

By plugging the above results into Eq. (3.4), (3.5), we obtain:

$$\mu_s^g \equiv \mathbb{E}^g[s] = 1 - \frac{1-a}{1+a}\int_0^1 dZ_1\, Z_1 \int_{Z_1}^{2-Z_1} dZ_2 \frac{1}{Z_2}\mathscr{N}^B_{1/2}\left(Z_1, Z_2; \boldsymbol{\mu}_{1,2}, \boldsymbol{\Sigma}_{1,2}\right) - \frac{2a}{1+a} \tag{3.30}$$

$$\begin{aligned}\left(\sigma^2\right)_s^g \equiv \mathrm{Var}^g(s) = \frac{1}{q}\left\{\frac{1-a}{1+a}\int_0^1 dZ_1 \int_{Z_1}^{2-Z_1} dZ_2 \left(\frac{Z_1}{Z_2} - \mathbb{E}^g[Z]\right)^2 \mathscr{N}^B_{1/2}\left(Z_1, Z_2; \boldsymbol{\mu}_{1,2}, \boldsymbol{\Sigma}_{1,2}\right)\right. \\ \left. + \frac{2a}{1+a}\frac{1}{\mathscr{Z}}\int_0^1 dZ_1 \left(1 - \mathbb{E}^g[Z]\right)^2 \mathscr{N}^B\left(Z_1, Z_1; \boldsymbol{\mu}_{w,w}, \boldsymbol{\Sigma}_{1,1}\right)\right\}\end{aligned} \tag{3.31}$$

The four formulas Eq. (3.17), (3.18), (3.30), (3.31) are the final result of the statistical analysis and they will be discussed in the Results section.

3.1.5 Model network with plastic weights

Below we describe the model neural network on which we will apply our symmetry measure.

Single-neuron dynamics We simulated $N = 30$ leaky integrate-and-fire neurons (Dayan and Abbott [2001]) with a firing threshold of $V_{thr} = -50\,mV$. The sub-threshold dynamics of the electrical potential V_i is given by:

$$\tau_m \frac{dV_i}{dt} = -\left(V_i(t) - V_{rest}\right) + RI_i(t)\,, \tag{3.32}$$

where τ_m is the membrane time constant, V_{rest} is the resting potential, R is the membrane resistance and $I_i(t)$ is the input signal. We chose $\tau_m = 10\,ms$, $V_{rest} = -70\,mV$, $R = 1\,K\Omega$. To introduce noise in the firing process of neurons, we implemented the escape noise model

Symbol	Description	Value
N	Number of neurons	30
τ_m	Membrane time constant	10 ms
R	Membrane resistance	1 KΩ
V_{rest}	Resting and after-spike reset potential	-70 mV
V_{thr}	Threshold potential for spike emission	-50 mV
α_{syn}	Voltage increase due to a presynaptic event	1 mV
α_{ext}	Voltage increase due to an external event	30 mV
w_{min}	Lower bound for synaptic weights	0
w_{max}	Higher bound for synaptic weights	1
μ_w	Mean value of Gaussian-distributed initial weights	0.5
σ_w	Variance of Gaussian-distributed initial weights	0.01
A_2^+	Amplitude of weights change - pair term in LTP	4.6×10^{-3}
A_3^+	Amplitude of weights change - triplet term in LTP	9.1×10^{-3}
A_2^-	Amplitude of weights change - pair term in LTD	3.0×10^{-3}
A_3^-	Amplitude of weights change - triplet term in LTD	7.5×10^{-9}
τ_{r_1}	Decay constant of presynaptic indicator r_1	16.8 ms
τ_{r_2}	Decay constant of presynaptic indicator r_2	575 ms
τ_{o_1}	Decay constant of postsynaptic indicator o_1	33.7 ms
τ_{o_2}	Decay constant of postsynaptic indicator o_2	47 ms
γ	Learning rate for STDP	$\{1,7\}$
dt	Discretisation time step	1 ms
N_{iter}	Number of independent repetitions of the experiment	50

Table 3.1 List of parameters used for the case study. STDP parameters are as in the nearest-spike triplet-model, described in Pfister and Gerstner [2006].

(Gerstner and Kistler [2002b]). At each time-step Δt the probability that the neuron i fires is given by:

$$\wp_{dt}^{f}(V_i) = 1 - \exp\left(-\Delta t \rho_0 \exp\left(\frac{V_i - V_{thr}}{\Delta V}\right)\right) \tag{3.33}$$

where $\rho_0 = 0.1\,ms^{-1}$ and $\Delta V = 5\,mV$. Once a neuron fires, its membrane potential is reset to the resting value.

Synaptic and external inputs The input $I_i(t)$ to each neuron has two components: A synaptic part, coming from the action potentials of the other neurons, and an external part, which is defined by the applied protocol:

$$I_i(t) = I_i^{syn}(t) + I_i^{ext}(t) = \alpha^{syn} \sum_{j \neq i} w_{ij} \delta\left(t - t_j^f - \varepsilon\right) + \alpha^{ext} \delta\left(t - t_i^{ext}\right). \tag{3.34}$$

In the synaptic term, w_{ij} are the synaptic weights, t_j^f is the firing time of the presynaptic neuron j and ε is a small positive number accounting for the delivering time of the electrical signal from the presynaptic to the postsynaptic neuron. The term t_i^{ext} is the time course of the injected input, which is different from neuron to neuron and depends on the protocol we use (see Results section). Finally, the amplitudes α^{syn} and α^{ext} are fixed to the same value for all neurons. We chose $\alpha^{syn} = 1\,\mu A \times s$ and $\alpha^{ext} = 30\,\mu A \times s$, so that each external input forces the neurons to fire.

Plasticity The efficacy of the synaptic connections is activity-dependent. Therefore, the unpruned elements of the adjacency matrix w_{ij} in Eq. (3.34) change in time by Spike-Timing Dependent Plasticity (STDP) mechanisms, i.e. passively driven by the input protocol and emerging internal dynamics, without the presence of a supervisory or reinforcement learning signal (Richmond et al. [2011]; Vasilaki et al. [2009a,b]). More specifically, we implemented the triplet STDP rule (Clopath et al. [2010, 2008]; Pfister and Gerstner [2006]) with parameters from Pfister and Gerstner [2006] (Visual cortex, nearest neighbour dataset), see Tab. 3.1, and we constrain the connections in $[0, 1]$. In this model, each neuron has two presynaptic variables r^1, r^2 and two postsynaptic variables o^1, o^2. In the absence of any activity, these variables exponentially decay towards zero with different time constants:

$$\tau_{r_1}\frac{dr_i^1}{dt} = -r_i^1 \quad \tau_{r_2}\frac{dr_i^2}{dt} = -r_i^2 \quad \tau_{o_1}\frac{do_i^1}{dt} = -o_i^1 \quad \tau_{o_2}\frac{do_i^2}{dt} = -o_i^2 \tag{3.35}$$

whereas when the neuron elicits a spike they increase by 1:

$$r_i^1 \to r_i^1 + 1 \quad r_i^2 \to r_i^2 + 1 \quad o_i^1 \to o_i^1 + 1 \quad o_i^2 \to o_i^2 + 1\,. \tag{3.36}$$

Then, assuming that neuron i fires a spike, the STDP implementation of the triplet rule can be written as follows:

$$\begin{cases} w_{ji} \to w_{ji} - \gamma o_j^1(t)\left[A_2^- + A_3^- r_i^2(t-\varepsilon)\right] \\ w_{ij} \to w_{ij} - \gamma r_j^1(t)\left[A_2^+ + A_3^+ o_i^2(t-\varepsilon)\right] \end{cases} \tag{3.37}$$

where γ is the learning rate and ε is an infinitesimal time constant to ensure that the values of r_i^2 and o_i^2 used are the ones right before the update due to the spike of neuron i. The learning rate used is 1 for the frequency protocol, 7 for the sequential protocol (see Results).

Reproducibility of results All simulations were performed in MATLAB (The Mathworks, Natick, USA). Code is available from ModelDB (Hines et al. [2004]), accession number: 151692.

3.2 Results

We recall the definition of the symmetry measure s (Eq. 3.1):

$$s = 1 - \frac{2}{N(N-1) - 2M} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{\left| w_{ij} - w_{ji} \right|}{w_{ij} + w_{ji}} , \tag{3.38}$$

where w_{ij} is the positive synaptic connection from neuron j to neuron i, N is the total number of neurons and M is the number of instances where both w_{ij} and w_{ji} are zero, i.e. there is no connection between two neurons. The term $q = \frac{N(N-1)-2M}{2}$ is a normalisation factor that represents the total number of synaptic connection pairs that have at least one non-zero connection.

By using this definition, we were able to estimate the expected value and the variance of s on random matrices (uniform and truncated Gaussian), see Eq. (3.17)-(3.18) and (3.30)-(3.31) correspondingly. This provides us a tool to estimate the significance of the "symmetry" or "asymmetry" of the adjacency matrix of a given network, shaped by learning, given the initial distribution of the synaptic connections prior to the learning process. The statistical analysis is particularly useful in cases where the developed configuration is not "clear-cut", i.e. all connections have been turned to either bidirectional or unidirectional resulting in a symmetry measure almost 1 or 0, which is probably an artificial scenario, but rather in the intermediate cases, where we need a measure of how far away the value of the symmetry measure of a specific configuration is from that of a random configuration. Though here we focused on two specific random distributions, our methodology is applicable to other distribution choices.

3.2.1 Hypothesis test

Having calculated the mean and variance of the symmetry measure s over random networks of a specific connectivity distribution, we are now able to directly evaluate the symmetry measure s_s of a specific connectivity structure and conclude whether the symmetric or asymmetric structure observed is due to chance or it is indeed significant. A simple test is, for instance, to calculate how many standard deviations s_s is away from μ_s. Equivalently, we

can form the hypothesis that the configuration s_s is non-random and calculate the p-value by:

$$p\text{-value} = \pm 2 \int_{s_s}^{\pm\infty} \mathcal{N}(\mu_s, \sigma_s)\, ds\,, \tag{3.39}$$

where we implicitly assume that the distribution of the symmetry measure s over all random networks is Gaussian. We can compare this result with the significance level we fixed, typically $p_s = 0.05$, and we can then conclude the nature of the symmetry of the network with a confidence level equal to p_s or reject the hypothesis.

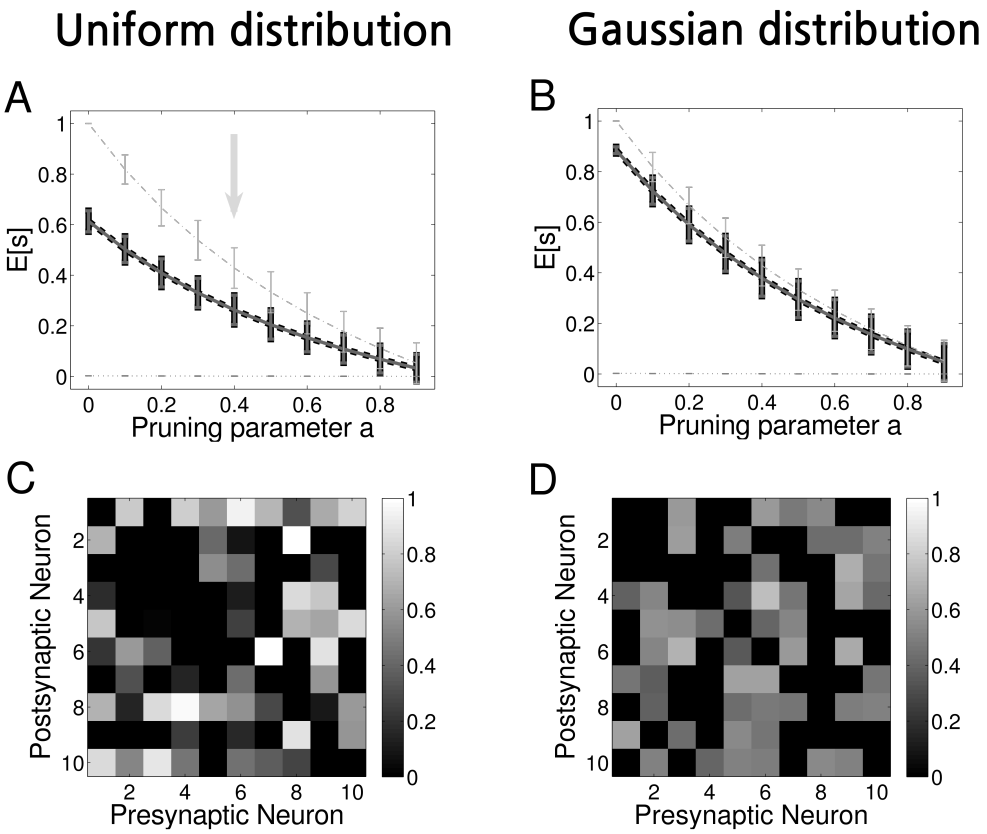

Figure 3.3 Final statistics of the symmetry measure. A Expected value and standard deviation of the symmetry measure as a function of the pruning for different types of networks with uniform weights distribution. The total length of each bar is two times the standard deviation. *Dashed light grey line*: Simulations for symmetric networks, *Dash dotted light grey line*: Simulations for asymmetric networks, *Solid dark grey line*: Simulations for random networks, *Dashed black line*: Theoretical results for random networks. **B** The same as A but with Gaussian-distributed random weights. **C** Example of an adjacency matrix in a particular random network with uniform weights distribution and pruning parameter $a = 0.4$. For this example $s \simeq 0.260$. **D** The same as C but with Gaussian-distributed random weights. For this example $s \simeq 0.322$.

a	$\mu_s^u \pm \sigma_s^u$	$\mu_s^g \pm \sigma_s^g$
0.0	0.614 ± 0.042	0.885 ± 0.013
0.1	0.502 ± 0.052	0.724 ± 0.053
0.2	0.409 ± 0.056	0.590 ± 0.064
0.3	0.331 ± 0.058	0.476 ± 0.070
0.4	0.263 ± 0.058	0.379 ± 0.072
0.5	0.205 ± 0.057	0.295 ± 0.072
0.6	0.153 ± 0.056	0.221 ± 0.072
0.7	0.108 ± 0.055	0.156 ± 0.071
0.8	0.068 ± 0.053	0.098 ± 0.070
0.9	0.032 ± 0.052	0.047 ± 0.068

Table 3.2 Mean value and standard deviation of the symmetry measure as obtained from the theoretical analysis. *Column 1:* Value of the pruning parameter a. *Column 2:* Uniform distribution. *Column 3:* Gaussian distribution. These values are obtained with $n = 10^5$ random networks of $N = 10$ neurons and are plotted in Fig. 3.3.

3.2.2 Pruning biases the network towards asymmetry

To demonstrate the validity of our analytical results, we compare them to simulation results. We generated a sample of $n = 10^5$ networks with $N = 10$ neurons with random connections with synaptic efficacies varying from 0 to 1. We evaluated the symmetry measure on each network by applying directly the definition of Eq. (3.38), and then we computed the mean value and variance of that sample. This process was repeated ten times, each one for a different value of the pruning parameter, $a = \{0, 0.1, 0.2, \ldots 0.9\}$. The final results are shown in Fig. 3.3A,B, together with the analytical results, see Eq. (3.17) and (3.30). Since numerical and analytical results overlap, we used a thicker (black) line for the latter. The agreement between theoretical findings, listed in Tab. 3.2, and numerical evaluations is excellent.

We also considered two extreme cases, symmetric and asymmetric random networks, which respectively represent the upper and lower bound for the symmetry measure defined in Eq. (3.38). Symmetric random networks have been generated as follows: We filled the upper triangular part of the $N \times N$ weights matrix with random values from the uniform/Gaussian distribution. We then mirrored the elements around the diagonal so as to have $w_{ij} = w_{ji}$. In the asymmetric case, instead, we generated a random adjacency matrix with values in $[0.1, 1]$ for the upper triangular part and in $10^{-3} \times [0.1, 1]$ for the lower triangular part, so as to have $w_{ij} \ll w_{ji}$. Then, we shuffled the adjacency matrix.

In Fig. 3.3A,B we contrast our results on random networks with numerical simulations of symmetric and asymmetric random networks: The dashed, light grey line (top line) shows the upper extreme case of a symmetric random network $w_{ij} = w_{ji}\ \forall i, j = 1 \ldots N$, whereas

the dash-dotted, light grey line (bottom line) shows the lower extreme case of a asymmetric random network $w_{ij} \ll w_{ji}$ for $i > j$.

When we introduce pruning, the lower bound of s remains unchanged, whereas the more we prune the more a symmetric network appears as asymmetric.

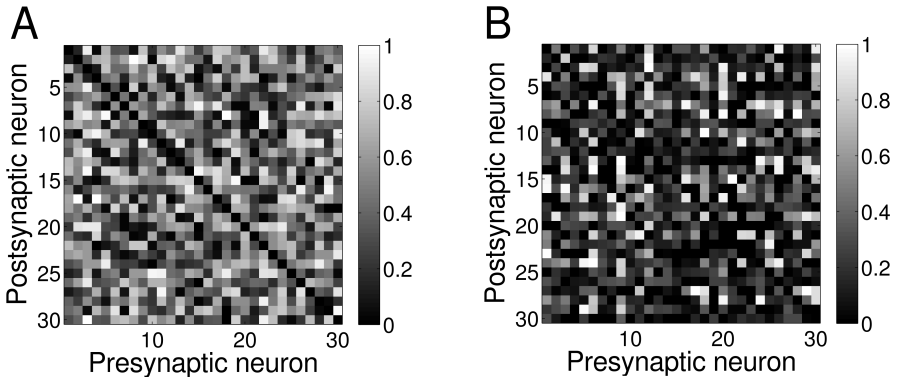

Figure 3.4 Symmetry and asymmetry depends on the distribution of the initial connectivity. A Example of an adjacency matrix in a random network with pruning parameter $a = 0$ and symmetry measure $s \simeq 0.900$. According with the p-value test with the null hypothesis of random connectivity and with a level of confidence of 0.05, the symmetry of this network is significant if the distribution of the initial connections is uniform but is non-significant if the initial distribution of the connections is Gaussian. Therefore, in the first case it should be regarded as a non-random network whereas in the second case as a random network. **B** The same as A but with pruning parameter $a = 0.2$ and symmetry measure $s \simeq 0.334$. In this case, with the same hypothesis test, the situation is reversed: The network should be considered random for initial uniform distribution of connections, but non-random for initial Gaussian-distributed connections (see the discussion in the text).

3.2.3 Gaussian-distributed synaptic efficacies bias the network towards symmetry

In Fig. 3.3C,D, we show the adjacency matrix W for a random pruned network with pruning parameter $a = 0.4$. A network with uniformly distributed initial connectivity is shown in Fig. 3.3C and a network with Gaussian-distributed initial connectivity is shown in Fig. 3.3D. Black areas represent zero connection, $w_{ij} = 0$. The "Gaussian" network has most of the connections close to the mean value μ_w, resulting in higher values for the symmetry measure than in the case of a uniform distribution, compare Fig. 3.3B with Fig. 3.3A.

This difference in the mean values of s depending on the shape of the distribution implies that for example a weight configuration that would be classified as non-random under the hypothesis that the initial connectivity, before learning, is uniform, is classified as random under the hypothesis that the initial distribution of the connections is Gaussian. To more

emphasise this point, we show in Fig. 3.4 the adjacency matrix of two different networks of 30 neurons. The first network, Fig. 3.4A, is a non-pruned network with $s \simeq 0.900$. According with the values obtained from the statistical analysis (Tab. 3.2), if we assume that the connections of this network are randomly drawn from a uniform distribution, the p-value test (Eq. (3.39)) gives us p-value $\simeq 6.50 \times 10^{-12}$. With the usual confidence level of $p_s = 0.05$, this is a significant result, implying that the network configuration is unlikely to be random. Conversely, if we assume that the initial connectivity is drawn from a Gaussian distribution, we obtain p-value $\simeq 0.25$, meaning that the network configuration should be considered random.

In Fig. 3.4B, we show a pruned network with $a = 0.2$ and $s \simeq 0.334$. In this case the opposite is true: Under the hypothesis of uniform random initial connectivity, the network should be considered random, as p-value $\simeq 0.18$. Under the hypothesis of Gaussian-distributed random initial connectivity, the network should be considered asymmetric, as p-value $\simeq 7.20 \times 10^{-5}$.

3.2.4 Relation between symmetry measure and motifs

In what follows, we demonstrate the relation between our symmetry measure and unidirectional and bidirectional motifs. From the definition s, Eq. (3.38), we can deduct that in the extreme case of a network with unidirectional motifs, i.e. pairs of the form (0,x), $x > 0$, the symmetry measure will result in $s = 0$, while in the case of bidirectional motifs i.e. pairs of the form (x,x), the symmetry measure will result in $s = 1$. By inverting Eq. (3.4), we can derive the mean value for connection pairs $\mu_Z \equiv \mathbb{E}_n[Z_k] = 1 - \mu_s$. We can use now this value to define connection pairs in a network as unidirectional or bidirectional: If $Z_k \geq \mu_Z$ than Z_k is a unidirectional motif, otherwise it is a bidirectional motif. In this way we relate unidirectional and bidirectional motifs to what is traditionally called single edge motif and second-order reciprocal motif, respectively. It is then expected that when s increases, the fraction of bidirectional motifs increases towards 1, whereas the percentage of unidirectional motifs decreases towards 0.

We show this relation in simulations by generating 10^3 networks of 15 neurons each, with uniformly distributed random connections in $[0, 1]$ and no pruning. In this case the mean value of the symmetry measure is $\mu_s^u \simeq 0.614$. Using Eq. (3.4), we have $\mu_Z^u \simeq 0.386$, which is the value used to decide whether a connection pair is unidirectional or bidirectional. For each of these networks, we calculated the value of the symmetry measure and the fraction of unidirectional and bidirectional motifs and we plotted the results in Fig. 3.5A as a scatter plot (black circles - bidirectional motifs, grey circles - unidirectional motifs). Also, un Fig. 3.5B

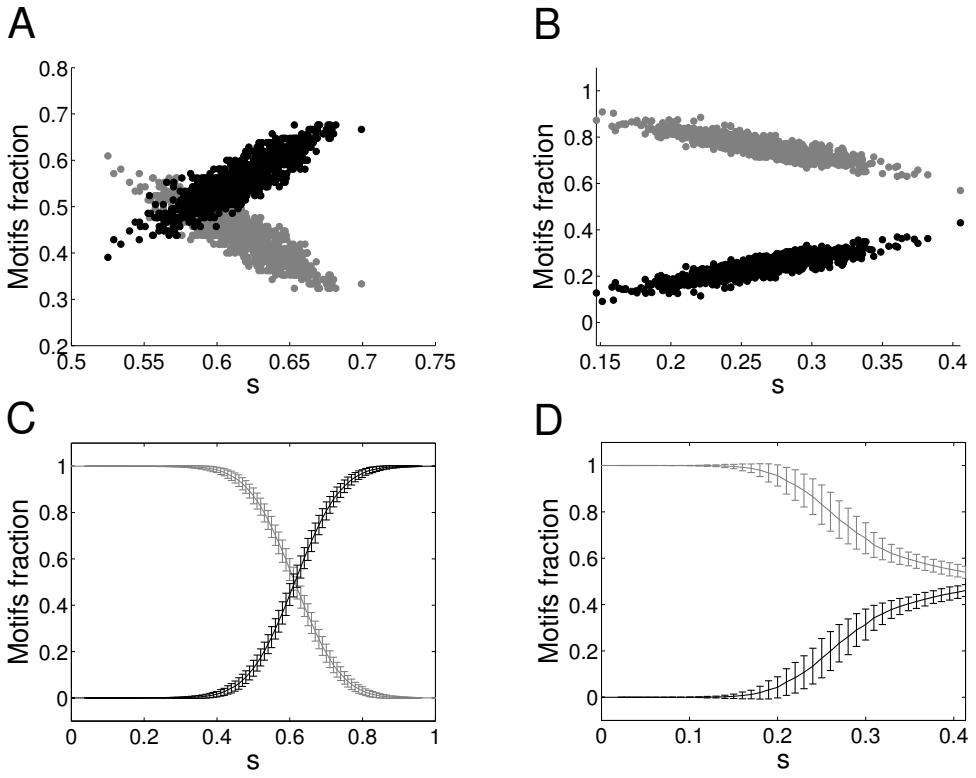

Figure 3.5 Symmetry measure reflects motifs formation. **A** Scatter plot of fraction of unidirectional and bidirectional motifs as a function of the symmetry measure for 10^3 networks with uniform random connections and $a = 0$. *Black dots*: Bidirectional motifs, *Grey dots*: Unidirectional motifs. For this typology $\mu_s^u \simeq 0.614$. **B** The same as A but with pruning parameter $a = 0.4$. In this case $\mu_s^u \simeq 0.263$. **C** Mean value and standard deviation (each bar is twice the standard deviation) of fraction of unidirectional and bidirectional motifs as a function of the symmetry measure for 10^3 networks with half of the connections uniformly distributed and $a = 0$. The second half of the connections were derived from the values of connection pairs Z, drawn from a Gaussian distribution with mean $1 - s$ and standard deviation 0.1. *Black line*: Bidirectional motifs, *Grey line*: Unidirectional motifs. **D** The same as C but with pruning parameter $a = 0.4$.

we show the analogous results obtained when we prune the connections with $a = 0.4$. In both cases, a linear relation between s and motifs is evident.

Note that in both figures the restricted domain on the s-axis: This is determined by the range of s values that correspond to random networks. If we want to extend this range, we need to consider networks that are not random any more. We achieve this by fixing a distribution for connection pairs Z. Once we decide on the desirable value of s, in our case the whole zero to one spectrum, we can use a distribution (e.g. Gaussian) with mean $\mu_Z = 1 - \mu_s$ and a chosen variance to draw the values of all the connection pairs in the network. Following this procedure, we fill the upper triangular part of the 15×15 weights matrix with random

values from the uniform/Gaussian distribution, and derive the other half of the weights by inverting the definition of Z. As a PDF(Z) we chose a Gaussian distribution around μ_Z with $\sigma = 0.1$, except for the extreme cases (near $s = 0$, $s = 1$) where $\sigma = 0$. With this technique of creating networks, we sampled the entire domain of s in steps of 0.01. For each value, we again generated 10^3 networks of 15 neurons with (half of the) weights uniformly distributed, and then we computed the mean value and standard deviation. Results are shown in Fig. 3.5C,D respectively for unpruned and pruned (with $a = 0.4$) networks (black line - bidirectional motifs, grey line - unidirectional motifs). We can see that Fig. 3.5C,D correctly reproduce the linear regime observed in Fig. 3.5A,B for values of s close enough to μ_s^u.

Due to the method by which we generated networks, the shape of the distribution of half of the weights does not affect the shape of the dependence in Fig. 3.5C,D. Indeed, if we choose half of the connections to be Gaussian-distributed, we will observe only a shift in both curves as they have to cross at $\mu_s^g = 0.885$ (results not shown).

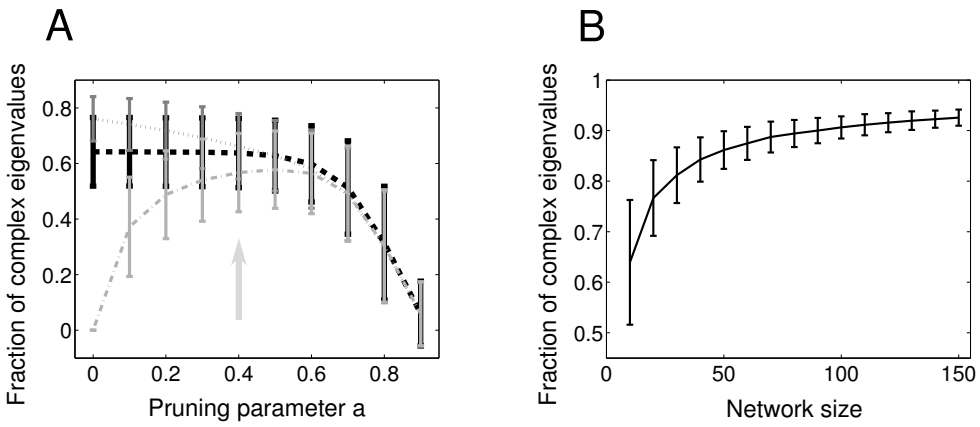

Figure 3.6 Eigenvalues and network structure. A Expected value and standard deviation of the fraction of complex eigenvalues as a function of the pruning for different types of networks of $N = 10$ neurons with uniform weight distribution. The total length of each bar is two times the standard deviation. *Dotted, dark grey line*: Simulations for asymmetric networks, *Dashed, black line*: Simulations for random networks, *Dash-dotted, light grey line*: Simulations for symmetric networks. **B** Fraction of complex eigenvalues as a function of network size for random networks with uniform weights distribution. Pruning parameter $a = 0$.

3.2.5 Symmetry measure and eigenvalues

In the definition of our symmetry measure we have deliberately excluded $(0,0)$ connection pairs. This was a conscious decision for mathematical and practical reasons, see Methods. As a consequence, pairs of the form $(0,0)$ do not contribute to the evaluation of the symmetry of

the network. Instead, pairs of the form $(0,\varepsilon)$, with ε very small, contribute to the asymmetry of the network according to our specific choice of symmetry measure (leading to $Z=1$, see Methods). Here we further motivate this choice via a comparison of our measure to the evaluation of the symmetry via the matrix eigenvalues, for three types of networks: (i) Symmetric, where each connection pair consists of synapses of the same value, (ii) asymmetric, where every connection pair has one connection set to a small value ε, and (iii) random, where connections are uniformly distributed. We demonstrate that our measure has a clear advantage over the eigenvalues method, in particular when pruning is introduced. This difference in performance lays in the different ways that $(0,0)$ and $(0,\varepsilon)$ are treated by our measure.

A crucial property of the real symmetric matrices is that all their eigenvalues are real. Fig. 3.6A depicts the fraction of complex eigenvalues vs the pruning parameter a for a symmetric (dash-dotted, light grey line) asymmetric (dotted, dark grey line) and random (dashed, black line) matrix with uniformly distributed values, similar to Fig. 3.3A, with the same statistics (10^5 networks of 10 neurons). As expected, if no pruning takes place ($a=0$), symmetric matrices have no complex eigenvalues and are clearly distinguishable from random and asymmetric matrices. On the contrary, both random and asymmetric matrices have a non-zero number of complex eigenvalues, which increases with a higher degree of asymmetry, leading to a considerable overlap between these two cases, differently from what happens with our measure in Fig. 3.3A.

As we introduce pruning, the mean of the complex eigenvalues of the three distinctive types of network moves towards the same value, an increase for the symmetric network and decrease for the random and non-symmetric networks. This is expected as pruning specific elements will make the symmetric network more asymmetric while it will increase the symmetry of the asymmetric network by introducing pairs of the form $(0,\varepsilon)$ or $(0,0)$. The $(0,\varepsilon)$ pairs are due to the construction of the asymmetric network, where half of the connections are stochastically set to very low values. This continues till $a=0.5$, after which further pruning reduces the number of complex eigenvalues of all networks: A high level of pruning implies the formation of more $(0,\varepsilon)$ or $(0,0)$ pairs for the asymmetric network and more $(0,0)$ pairs for the symmetric network. In Fig. 3.6B we show the dependence of the fraction of complex eigenvalues for uniform random matrices on their size.

Comparing Fig. 3.6A to Fig. 3.3A, we observe that our symmetry measure offers excellent discrimination between the symmetric, asymmetric and random matrices for e.g. $a=0.4$. This is despite the fact that the structure of the asymmetric matrix *per se* has become less asymmetric and the structure of the symmetric matrix has become more asymmetric due to the pruning, as it is confirmed by the overlapping fraction of complex eigenvalues for asymmetric

and random matrices (Fig. 3.6A). In our measure $(0,\varepsilon)$ pairs are treated as asymmetric, $(0,0)$ pairs are ignored, and the bias that pruning introduces is taken into account allowing for good discrimination for all types of matrices, even beyond $a = 0.4$.

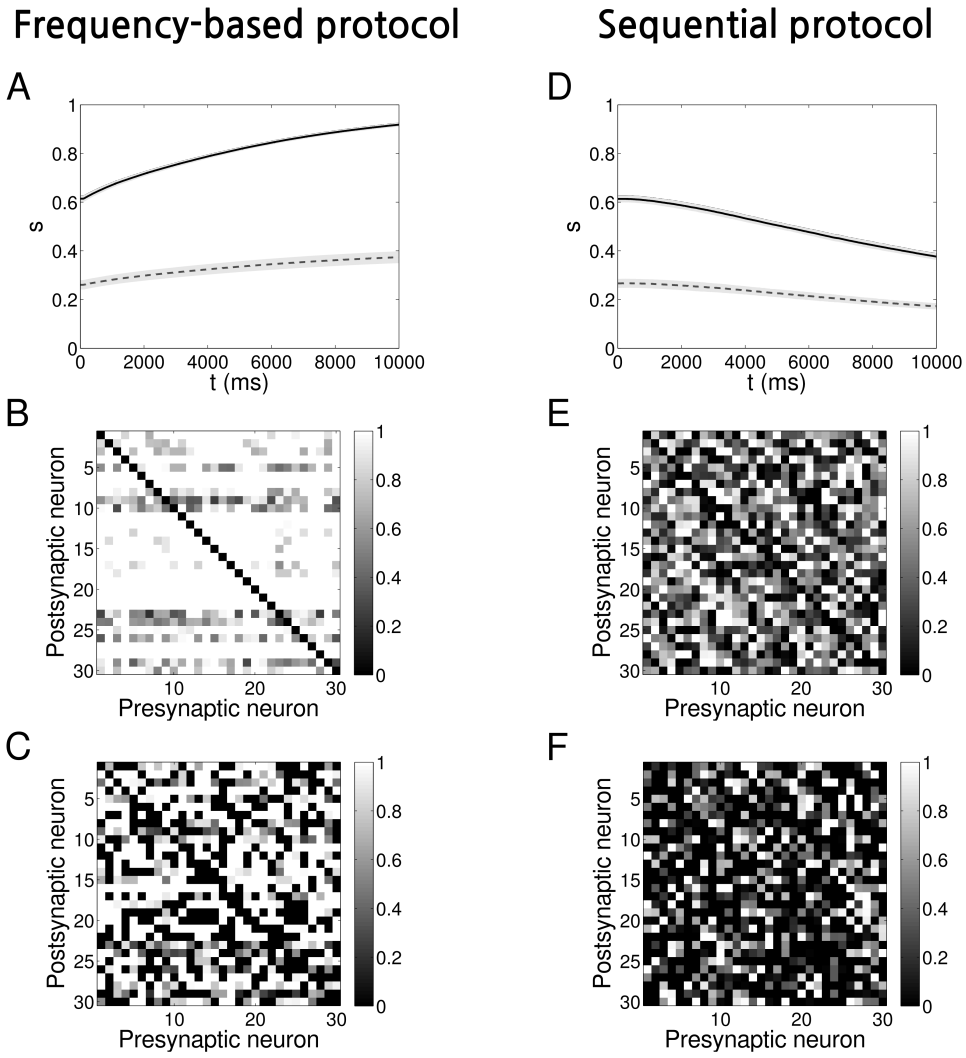

Figure 3.7 Evolution of networks with STDP and initially uniform weights distribution. A Time evolution of the symmetry measure when a frequency protocol is applied on a network, shown as average over 50 representations. The shaded light grey areas represent the standard deviation (the total length of height of each band is twice the standard deviation). *Solid black line*: No pruning, *Dashed grey line*: With pruning $a = 0.4$. **B** Example of an adjacency matrix at the end of the learning process for a network with the frequency protocol and no pruning. For this example $s \simeq 0.921$. **C** The same as B but with pruning $a = 0.4$. For this example $s \simeq 0.427$. **D - F** The same as A - C but with the sequential protocol applied. The connectivity matrix in panel E has $s \simeq 0.393$. The connectivity matrix in panel F has $s \simeq 0.141$.

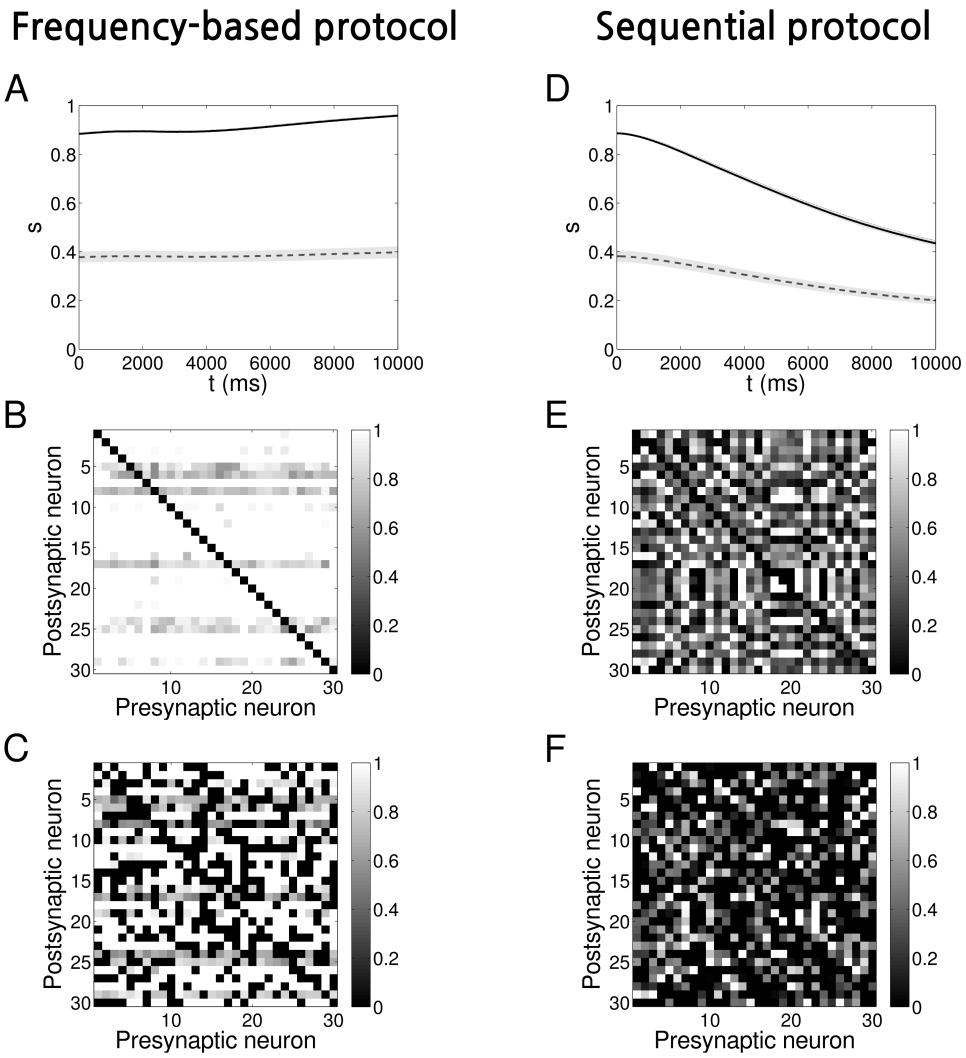

Figure 3.8 Evolution of networks with STDP and initially Gaussian-distributed weights. A Time evolution of the symmetry measure when a frequency protocol is applied on a network, shown as average over 50 representations. The shaded light grey areas represent the standard deviation (the total length of height of each band is twice the standard deviation). *Solid black line*: No pruning, *Dashed grey line*: With pruning $a = 0.4$. **B** Example of an adjacency matrix at the end of the evolution for a network with frequency protocol and no pruning. For this example $s \simeq 0.963$. **C** The same as B but with pruning $a = 0.4$. For this example $s \simeq 0.456$. **D - F** The same as A - C but with the sequential protocol applied. The connectivity matrix in panel E has $s \simeq 0.426$. The connectivity matrix in panel F has $s \simeq 0.153$.

3.2.6 Case study: Monitoring the connectivity evolution in neural networks

We demonstrate the application of the symmetry measure to a network of neurons evolving in time according to a Spike-Timing Dependent Plasticity (STDP) "triplet rule" (Pfister and Gerstner [2006]) by adopting the protocols of Clopath et al. [2010]. These protocols are designed to evolve a network with connections modified according to the "triplet rule", to either a unidirectional configuration or bidirectional configuration, with the weights being stable under the presence of hard bounds. We have deliberately chosen a small size network as a "toy-model" that will allow for visual inspection and characterisation at the mesoscopic scale.

We simulated $N = 30$ integrate-and-fire neurons (see Methods section for simulation details) initially connected with random weights $w_{ij} \in [0, 1]$ drawn from either a uniform (Fig. 3.7) or a Gaussian (Fig. 3.8) distribution (see Tab. 3.1 for parameters). Where a pruning parameter is mentioned, the pruning took place prior to the learning procedure: With a fixed probability some connections were set to zero and were not allowed to grow during the simulation.

Our choice allows us to produce an asymmetric or a symmetric network depending on the external stimulation protocol applied to the network. Since the amplitude of the external stimulation we chose ($\alpha^{ext} = 30mV$) is large enough to make a neuron fire every time it is presented with an input, the firing pattern of neurons reflects the input pattern and we can indifferently refer to one or another. The asymmetric network has been obtained by using a "sequential protocol", in which neurons fire with the same frequency in a precise order one after the other, with $5ms$ delay, see also Clopath et al. [2010]. The symmetric network is produced by applying a "frequency protocol", in which each neuron fires with a different frequency from the values $\{15, 16, 17, \ldots, 44Hz\}$. In both cases, the input signals were jittered in time randomly with zero mean and standard deviation equal to 2 % of the period of the input itself for the frequency protocol, to 25 % of the delay for the sequential protocol. Depending on the protocol, we expect the neurons to form mostly unidirectional or bidirectional connections during the evolution.

The time evolution for both protocols and initial distributions is shown in Fig. 3.7A,D (uniform) and 3.8A,D (Gaussian). Each panel represents the evolution of the symmetry measure averaged over 50 different representations for both fully connected networks ($a = 0$, solid black line) and pruned networks (e.g. $a = 0.4$, dashed grey line). The shaded area represents the standard deviation. The time course of the symmetry measure can be better understood with the help of the Fig. 3.3. At the beginning, the values of s reflect what we expect from a random network. Afterwards, as the time passes, the learning process leads to

the evolution of the connectivity. As expected, the frequency protocol induces the formation of mostly bidirectional connections, leading to the saturation of s towards its maximum value, depending on the degree of pruning. Conversely, when we apply the sequential protocol, connection pairs develop a high degree of asymmetry, the values of s decreasing towards its minimum. Connections were constrained to remain inside the interval $[0, 1]$.

Type	s	$\mu_s^{rand} \pm \sigma_s^{rand}$	p-value	$\mu_s^{s/a} \pm \sigma_s^{s/a}$
UFa_0	0.921	0.614 ± 0.042	1.82×10^{-13}	1.000 ± 0.000
UF$a_{0.4}$	0.427	0.263 ± 0.058	4.50×10^{-3}	0.429 ± 0.081
USa_0	0.393	0.614 ± 0.042	1.21×10^{-7}	$0.003 \pm 0.387 \times 10^{-3}$
US$a_{0.4}$	0.141	0.263 ± 0.058	3.46×10^{-2}	$0.001 \pm 0.352 \times 10^{-3}$
GFa_0	0.963	0.885 ± 0.013	3.30×10^{-9}	1.000 ± 0.000
GF$a_{0.4}$	0.456	0.379 ± 0.072	2.85×10^{-1}	0.429 ± 0.081
GSa_0	0.426	0.885 ± 0.013	0	$0.002 \pm 0.256 \times 10^{-3}$
GS$a_{0.4}$	0.153	0.379 ± 0.072	1.70×10^{-3}	$0.001 \pm 0.260 \times 10^{-3}$

Table 3.3 Symmetry measure and p-value for different types of network. *Column 1:* Network type. U = Uniform distribution, G = Gaussian distribution, F = Frequency protocol, S = sequential protocol, a_0 = No prune, $a_{0.4}$ = pruning of 0.4. *Column 2:* Value of the symmetry measure for one instance of each type. *Column 3:* Results from the previous statistical analysis on random networks. *Column 4:* Corresponding p-value from Eq. (3.39). *Column 5:* Results from the previous statistical analysis for the corresponding closest extreme case – symmetric network for frequency protocol and asymmetric network for sequential protocol. s means symmetric and a asymmetric.

The final connectivity pattern can be inspected by plotting the adjacency matrix W. In Fig. 3.7B,C and 3.8B,C we give an example of W at the end of the evolution for one particular instance of the 50 networks when the frequency protocol is applied. Similarly, in Fig. 3.7E,F and 3.8E,F we show the results for the sequential protocol. The corresponding values of s for each of the examples in the figures are listed in Tab. 3.3. In the case that $a = 0$, a careful inspection of Fig. 3.7B and 3.8B indicates that connectivity is bidirectional: All-to-all strong connections have been formed. Instead, in Fig. 3.7E and 3.8E, trying to determine if there is a particular connectivity emerging in the network starts to be considerably tough. However, by using our symmetry measure (see values in Tab. 3.3) we can infer that the connectivity is unidirectional. In the pruned networks, however, see Fig. 3.7C, 3.8C and Fig. 3.7F, 3.8F, the formation of bidirectional and unidirectional connection pairs is not as obvious as for $a = 0$. We therefore refer again to the Tab. 3.3 and compare the values of s with μ_s^{rand} and with μ_s^{asym} or μ_s^{sym}, depending on the case. We can then verify that the learning process has significantly changed the network and its inner connections from the initial random state.

We can rigorously verify the above conclusions via a statistical hypothesis test such as the p-value test, which in essence quantifies how far away the value of our symmetry measure s of our final configuration is from the initial, random configuration (see also Methods). In Tab. 3.3 we show the p-values corresponding to the null hypothesis of random connectivity for the examples in the Fig. 3.7, 3.8. Once we set the significance level at $p_s = 0.05$, we can verify that, except for the case of pruned network with initially Gaussian-distributed connections where a frequency protocol has been applied (i.e. GF$a_{0.4}$), the p-values are significant, implying the rejection of the null hypothesis. This is also justified by Fig. 3.3A,B: When we increase the pruning, the mean value of the symmetry measure of the fully symmetric network approaches that of the pruned random network and in particular for the case where the weight are randomly Gaussian-distributed.

3.3 Conclusion

The study of the human brain reveals that neurons sharing the same cognitive functions or coding tend to form clusters, which appear to be characterised by the formation of specific connectivity patterns, called motifs. We, therefore, introduced a mathematical tool, a symmetry measure s, which computes the mean value of the connection pairs in a network, and allows us to monitor the evolution of the network structure due to the synaptic dynamics. In this context, we applied it to a number of evolving networks with plastic connections that are modified according to a learning rule. After the network connectivity reaches a steady state as a consequence of the learning process, connectivity patterns develop. The use of the symmetry measure together with the statistical analysis and the p-value test allow us both to quantify the connectivity structure of the network, which has changed due to the learning process, and to observe its development. It also allows for some interesting observations. (i) Introducing a fixed amount of pruning in the network prior to the learning process biases the adjacency matrix towards an asymmetric configuration. (ii) A network configuration that appears to be symmetric under the assumption of a uniform initial distribution is random under the assumption of a Gaussian initial distribution.

Statements on non-random connectivity in motifs experimental work, e.g. Song et al. [2005]; Vasilaki and Giugliano [2014], are supported by calculating the probability of connectivity in a random network and then distributing it uniformly: This becomes the null hypothesis. This was a most suitable approach given the paucity of data. If, however, the null hypothesis consisted of a Gaussian-distributed connectivity, then a higher number of bidirectional connections would be expected, as suggested by our analysis.

Summary The measure introduced here applies to the adjacency matrix of an entire network and as such is a global estimator of the symmetry of the connectivity of a network. However, it is possible that in large networks learning processes are only modifying a subset of the connections, forming motifs that might be unobserved if the symmetry measure is applied to the whole matrix. Clearly, it can also be applied to a limited sub-region within the network, but without a clear *a priori* knowledge or indication on where to look for symmetric or asymmetric areas, this is not a meaningful procedure. In such cases, algorithms are needed to detect potential symmetric or asymmetric regions. The symmetry measure presented here could be therefore applied to such areas only to reveal the evolution of the structure and its significance. Next chapter addresses this question by defining a heuristic algorithm for a particular class of structures.

Chapter 4

Detection of Multiple and Overlapping Bidirectional Communities within Large, Directed and Weighted Networks of Neurons[1]

When studying motifs formation in networks of neurons, it is most likely that learning affects not the entire network but only subsets of neurons, which are not known in advance and are very difficult to detect. A global estimator like the one defined in the previous chapter is therefore no longer meaningful in these cases.

Graph and Network Theory represent the natural ground to tackle problems related to structures' detection, and indeed they have been recently applied to brain networks (He and Evans [2010]; Sporns [2011a, 2013]), producing good results (Bassett and Bullmore [2009]; Guye et al. [2010]). However, typical Network Theory problems related to structure searching are proven to be either NP-complete or NP-hard (Bomze et al. [1999]; Cook [1971]; Garey and Johnson [1990]; Papadimitriou [1977]; Wegener [2005]). It is therefore common procedure to employ heuristic algorithms. The purpose of the work presented here is to add a contribution to structures' detection in networks of neurons direction by means of a heuristic algorithm designed to identify a particular class of such structures.

Besides the computational limitations, networks of neurons are arguably the most challenging type of graph to deal with, as they are instances of directed and weighted graphs with continuous weights. Most studied problems in Network Theory are based on undirected networks (Fortunato [2010]), with some of them focussing on directed un-weighted (or

[1] This chapter is almost entirely taken from Esposito and Vasilaki [2015]

binary) graphs (Malliaros and Vazirgiannis [2013]). In most of the cases generalisation to directed and weighted graphs is not always trivial. Moreover, in general, there is no limitation on the number of structures that can be formed within a network of neurons, nor on their shapes and overlaps. This leads to a very generic problem that needs to be narrowed to design an effective searching algorithm.

On the other side, we show that having a network of neurons and structures that arise from learning allows us to make considerations and hypotheses that greatly simplify the searching task, ultimately framing it within the domain of community detection in Network Theory (Girvan and Newman [2002]; Newman [2004]). This field has received constantly increasing attention due to the fact that community structures are often present in many types of networks and through their study the understanding of the network itself can be greatly improved (Porter et al. [2009]). However, despite huge efforts of a large interdisciplinary community of scientists, the problem is not yet satisfactorily solved.

Most of the existing algorithms for community structure use techniques like hierarchical clustering (Girvan and Newman [2002]; Newman and Girvan [2004]), modularity optimisation (Danon et al. [2005]; Newman [2006]; Ovelgönne and Geyer-schulz [2012]), which is also a NP-complete problem (Brandes et al. [2006]), spectral searching (Newman [2013]) and statistical inference (Ball et al. [2011]; Rosvall and Bergstrom [2007]). These methods are usually not designed for directed and weighted networks and also they do not consider overlapping communities. Furthermore, each class has its own limitations. For instance, modularity optimisation, which is the most widely used method, is known to have resolution problems (Fortunato and Barthélemy [2007]), and spectral analysis is much more complex for directed graphs as it is characterized by asymmetrical matrices. Developing methods of community detection for directed graphs is a hard task. The most important class of algorithms for the complete problem, i.e. detection of overlapping communities in directed and weighted graphs, is the clique percolation method (Karrer et al. [2014]; Palla et al. [2005]). However, since it does not look for actual communities but just for regions containing many cliques, it fails in several scenarios and its success also depends on the quantity of cliques that are present in the network (Fortunato [2010]). Community detection also suffers from the lack of a unique definition: How to identify a community generally varies depending on the problem and on the algorithm, and often a community is just the final outcome of the algorithm itself (*a posteriori* approach) (Fortunato [2010]).

Here we start by giving a general definition of community (*a priori* approach) and we show how this represents a great advantage as such a definition can be used as a guidance for building the algorithm. Our method, which aims at detecting multiple and overlapping bidirectional communities in directed and weighted networks of neurons, is based on a

statistical analysis of connections and it is a mixture of different techniques. At the basis of the algorithm there is the notion of symmetry measure introduced by Esposito et al. [2014] as an indicator of the global symmetry of a network's connectivity. Below, we introduce a local version of this measure, which, together with the community definition, allows us to develop a peculiar searching technique, a mixture between top-down and bottom-up approaches, that does not require to look at single connections to identify communities. This first part already provides very good results and in a very short time, but is able to detect only the non-overlapping parts of communities . Following this, we implement a neuron by neuron evaluation, that we call friendship algorithm, where we restore the detailed information about which pairs of neuron are connected to each other. This greatly increases the total computational time but it also improves the accuracy on the final outcome and allows to detect overlapping regions as part of more than one community.

4.1 Methods

Consider a directed and weighted network of N nodes that are all-to-all connected, with connectivity matrix W. Without losing generality, we allow single connections w_{ij} to vary in $[0,1]$, where w_{ij} represents the strength of the connection from node j to node i. We do not consider self-interactions, i.e. $w_{ii} = 0 \; \forall i = 1 \ldots N$.

4.1.1 Preliminary assumptions

We assume that the network described by W is the result of some learning process that affects only (unknown) parts of the network, significantly shaping these connections away from their initial configuration. Hence:

- Prior to learning, there is no way to differentiate the neurons that are going to be affected by the process from the rest of the network. We therefore assume that before learning all the connections in the network are randomly drawn from the same distribution.
- Connections between any pair of neurons are subject to the same learning process and therefore evolve in a similar way, which constraints structures to have the particular shape of a blob, rather than, for example, of a filament or a ring.

As a result, structures appear as regular bumps that stand out of the global randomness of the network's connectivity. In addition, we take into account that learning can occur with an efficacy $\varphi \leq 100\%$ (some connections may be faulty and not evolve) and that it can be

slower for some neurons and faster for others. This makes the final blobs' connectivity far from being a perfect and regular structure: Locally, some connections may not display any feature of the learning process, but the majority of the connections in the structure does, which preserves the global property of forming a bump in the network's connectivity.

In what follows, we adopt and generalise the terminology from Network Theory and refer to these structures as communities.

4.1.2 Definition of community and bidirectional community

For unweighted graphs, a community is generally a region where the edge concentration inside is higher than outside (Fortunato [2010]). In this case, the community detection problem for a complete network has the network itself as the only, trivial solution. When searching for communities in continuous weighted networks, it is essential to specify with respect to which property of the connectivity we are investigating the community structure: By changing the feature of the connectivity we look at, a different community structure can emerge. We can therefore phrase the concept of community in terms of an over-expression, in this case, of some property related to the connectivity. Thus, a complete weighted graph, differently from an unweighted one, can present solutions different from the trivial ones and in principle it can offer the same variety in the community structure as a sparse unweighted network.

In our case, the feature we investigate is bidirectionality. Two neurons i, j form a bidirectional pair when both connections have a similar strength, $w_{ij} \simeq w_{ji}$, resulting in information flowing nearly equally in both directions. Guided by experimental results showing excess of bidirectional connections in some regions of the brain (Song et al. [2005]; Wang et al. [2006]), we assume that learning strengthens all the connections involved, thus acting as a Hebbian-like process (Clopath et al. [2010]; Esposito et al. [2015]; Richmond et al. [2011]; Vasilaki et al. [2009a]; Vasilaki and Giugliano [2012, 2014]). This leads to the formation of what we call bidirectional communities within the network: Subsets of neurons that show an over-expression of bidirectional connections among them, when compared with the rest of the network. Since connections are continuous variables, there is no clear way to discriminate a pair that is bidirectional from a pair that is not without using the threshold concept. In the following, we will describe how to fix a bidirectionality threshold, essential for the algorithm implementation.

Local estimator of bidirectionality Esposito et al. (Esposito et al. [2014]) introduced a measure of network's connectivity, ranging from 0 to 1, that for fully connected networks

reduces to the following:

$$s = 1 - \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{|w_{ij} - w_{ji}|}{w_{ij} + w_{ji}}. \tag{4.1}$$

The extreme values $s = 1$ and $s = 0$ respectively correspond to completely symmetric networks, for which $w_{ij} = w_{ji}\ \forall i, j = 1 \ldots N$, and to completely asymmetric networks, for which $w_{ij} = 0$, $w_{ji} \neq 0\ \forall i, j = 1 \ldots N$ with $i > j$. In between these extremes, there is a continuum of values with smooth transitions between bidirectional, random and unidirectional networks. Through a statistical analysis of this symmetry measure on random networks, it is possible to identify a *bidirectionality threshold* s_B, which depends on the distribution of connections, separating bidirectional networks from non bidirectional ones (Esposito et al. [2014]).

This is, however, a global indicator that cannot capture a deeper organisation at a sub-networks level, nor it can be directly used to find communities, as it would require an extensive search. However, it can be used *i)* to validate community candidates after a successful searching and *ii)* to construct a local estimator encoding for the bidirectionality feature. Indeed, the symmetry measure is a global average of a local pairwise quantity, the relative strength of a pair of connections, defined as:

$$Z_{ij} = \frac{|w_{ij} - w_{ji}|}{w_{ij} + w_{ji}} \tag{4.2}$$

Z is a continuous variable ranging from 0 to 1 that covers all the possible states in which a connection pair can be found. In particular, bidirectionality is expressed by $Z \to 0$. Similirarly to s, we can map this continuum into a discrete two-state space, corresponding to randomness and bidirectionality, by fixing a *local bidirectionality threshold* Z_B on the connection pair. This can be done by simply translating s_B into the corresponding value of Z by using the definition of s itself:

$$Z_B = 1 - s_B \tag{4.3}$$

This follows from the consideration that a network with all equal values of $Z_{ij} = \bar{Z}$, for which $\bar{s} = 1 - \bar{Z}$, must show the same property, for instance bidirectionality, both locally in each pair and globally.

Thus, a bidirectional community of neurons is a set of neurons within which the majority of all possible connection pairs satisfy the relation $Z \geq Z_B$, i.e. they are bidirectional.

Over-density indicator The loose concept of majority reflects the over-density property and it can be mathematically formalised by setting a *community threshold* $\vartheta_{\mathscr{C}} \leq 100\%$: A set of neurons is a bidirectional community when, for each neuron within it, at least $\vartheta_{\mathscr{C}}$ of the available connections with other the neurons in the set is bidirectional. This threshold is clearly related to the learning efficacy φ. For all-to-all connected networks, like the ones we are considering here, we can give the following formal definition:

Definition of Community Be $\mathscr{C}$ a set of neurons, $i \in \mathscr{C}$ a neuron and

$$\mathscr{S}^i_{\mathscr{C}} := \left\{ j \in \mathscr{C} : Z_{ij} \geq Z_B \right\} \tag{4.4}$$

the set in $\mathscr{C}$ of all and only the neurons that form a bidirectional pair with neuron i. Then $\mathscr{C}$ is a community with respect to the property of bidirectionality if and only if each neuron in $\mathscr{C}$ forms within $\mathscr{C}$ itself at least a number of bidirectional connections equal to a fraction $\vartheta_{\mathscr{C}}$ of the total available connections in $\mathscr{C}$. In formal terms, this can be expressed with the following set of equations:

$$|\mathscr{S}^i_{\mathscr{C}}| \geq \vartheta_{\mathscr{C}} \left(|\mathscr{C}| - 1 \right) \quad \forall i \in \mathscr{C} \,, \tag{4.5}$$

where $|\cdot|$ represents the cardinality of a set. Hence, $|\mathscr{S}^i_{\mathscr{C}}|$ is the number of connections between neuron i and the other neurons in $\mathscr{C}$ that are bidirectional, and $|\mathscr{C}| - 1$ represents the number of neurons in the community available for forming a bidirectional pair. The maximum value $\vartheta_{\mathscr{C}} = 100\%$ corresponds to the specific case of a clique, the final result of a perfectly efficient learning $\varphi = 100\%$. Eq. (4.5) clearly captures the main difficulty of the community detection problem: We want to find a set of neurons $\mathscr{C}$ whose definition relies on the sets $\left\{ \mathscr{S}^i_{\mathscr{C}} \right\}$, which in turn are defined in terms of $\mathscr{C}$ itself and are unknown, with the set $\left\{ |\mathscr{S}^i_{\mathscr{C}}| \right\}$ also being unknown.

4.1.3 Algorithm description

The algorithm we describe below aims at identifying multiple and overlapping bidirectional communities, as defined in Eq. (4.5), within large networks of neurons. This is achieved by a popularity ranking (*Step 1* below) followed by two different techniques that are applied in sequence (*Step 2* and *Step 3*). If implemented alone, each of them already offers good results, but the combination refines the search and in some cases it also makes it faster.

In Fig. 4.1 - 4.3 we show the algorithm implementation on a toy network of $N = 11$ neurons, all-to-all connected and labelled as $N_1, N_2, \ldots N_{11}$ (Fig. 4.1A, *left*). For simplicity, instead of using N_i when referring to the neurons of the example, we assume that indices like

Symbol	Description
N	Size of the network
w_{ij}	Strength of the single connection from neuron j to neuron i
Z_{ij}	Relative strength of the connection pair between neurons i and j
Z_B	Bidirectionality threshold for connection pairs
n^i	Number of bidirectional pairs formed by neuron i in the entire network
$\mathscr{P}$	Bidirectional pool
$N_{\mathscr{P}}$	Size of the bidirectional pool
$n^{min}_{\mathscr{P}}$	Minimum number of bidirectional pairs to be part of the pool
$n^i_{\mathscr{P}}$	Number of bidirectional pairs formed by neuron i within the pool
$\mathscr{C}$	Bidirectional community
$n^i_{\mathscr{C}}$	Number of bidirectional pairs formed by neuron i within the community
$\vartheta_{\mathscr{C}}$	Threshold for belonging to a community
$\mathscr{C}^{max}_i$	Largest possible community that neuron i can form in the pool
$\tilde{\mathscr{B}}$	Bidirectional candidate blob
$N_{\tilde{\mathscr{B}}}$	Size of the bidirectional candidate blob
$N^{max}_{\tilde{\mathscr{B}}}$	Size of the largest community the bidirectional candidate blob can be part of
$n^i_{\tilde{\mathscr{B}}}$	Number of bidirectional pairs formed by neuron i within the candidate blob
$n^{min}_{\tilde{\mathscr{B}}}$	Minimum number of bidirectional pairs that each neuron in the candidate blob needs to form within it
$\mathscr{B}$	Bidirectional blob
$n^i_{\mathscr{B}}$	Number of bidirectional pairs formed by neuron i within the blob
$\tilde{\mathscr{C}}$	Bidirectional candidate community
$N_{\tilde{\mathscr{C}}}$	Size of the bidirectional candidate community
$n^{min}_{\tilde{\mathscr{C}}}$	Minimum number of bidirectional pairs that a candidate neuron needs to form with the candidate community
$n^i_{\tilde{\mathscr{C}}}$	Number of bidirectional pairs formed by neuron i with the members of the current candidate community
ϑ_{noise}	Threshold for noisy communities
ϑ_{ω}	Threshold for communities merger

Table 4.1 List of the symbols used for the algorithm description and their meaning.

i vary directly in the set $N_1, N_2, \ldots N_{11}$. Moreover, for a better understanding, in Tab. 4.1 we report a list of the symbols used and their description.

Step 1. Neurons popularity ranking From the full network's connectivity W (Fig. 4.1B, *left*), we derived the relative strength Z_{ij} of each pair of neurons (Fig. 4.1B, *middle*), given by Eq. (4.2), and we assess their bidirectionality using the threshold Z_B defined in Eq. (4.3). This allows to assign to each neuron i a number n^i representing how many bidirectional pairs that neuron forms in the entire network (Fig. 4.1B, *right*).

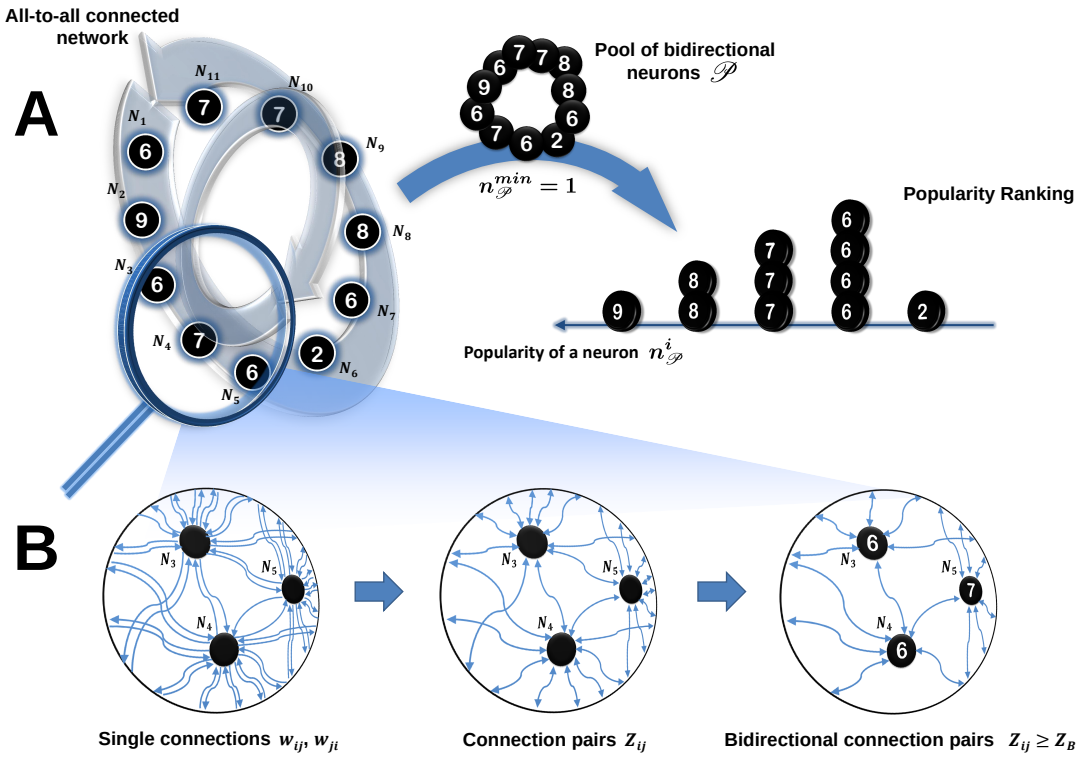

Figure 4.1 Algorithm Step 1: Neurons popularity ranking. **A** All-to-all toy network of $N = 11$ neurons labelled from N_1 to N_{11} (*left*). Each neuron i is associated with an integer representing the number of bidirectional connections made by i in the entire network, n^i. Neurons meeting the threshold $n^{min}_{\mathscr{P}} = 1$ are entered in a bidirectional pool $\mathscr{P}$ (*middle*), and here they are sorted in a popularity ranking according to the number of bidirectional connections made within the pool, $n^i_{\mathscr{P}}$ (*right*). In this example, network and pool coincide. **B** Zoom on a portion of the network highlighting the procedure to obtain $\{n^i\}$: The initial directed network $\{w_{ij}, w_{ji}\}$ (*left*) is mapped into an undirected network of single connection pairs $\{Z_{ij}\}$ (*middle*). n^i counts how many of these pairs fall in the bidirectional domain (*right*). This quantity is used as initial criterion to enter the neurons in the pool $\mathscr{P}$ (see text).

Based on this information, neurons are initially entered in a bidirectional pool $\mathscr{P}$ depending on a minimum required number of bidirectional pairs $n^{min}_{\mathscr{P}}$, that can be arbitrarily chosen (Fig. 4.1A, *middle*). Neurons that do not meet the *pool entering condition* $n^i \geq n^{min}_{\mathscr{P}}$ are excluded from $\mathscr{P}$, as they do not have the basic requirement for being part of a community. As a consequence, the bidirectional pairs that these excluded neurons form in the network also cannot be part of any community, hence they should be subtracted from the $\{n^i\}$ of the involved neurons. Therefore, after $\mathscr{P}$ is formed, neurons are subject to the *pool staying condition* $n^i_{\mathscr{P}} \geq n^{min}_{\mathscr{P}}$, where $n^i_{\mathscr{P}}$ is the number of bidirectional pairs formed only within $\mathscr{P}$. Nodes violating this inequality are excluded from $\mathscr{P}$ and so are their bidirectional pairs. The pool is therefore reduced and $\{n^i_{\mathscr{P}}\}$ need to be updated. This iterative process stops

when $n^i_{\mathscr{P}} \geq n^{min}_{\mathscr{P}}\ \forall i \in \mathscr{P}$ or when the number of neurons left in the pool is below the noise threshold (meaning that an eventual community can be considered as a random happening, see below). In the first case, the final $\mathscr{P}$ is the working material for the next steps, whereas in the second case the entire algorithm ends with no communities found.

Differently from the following steps, nodes that are left out of $\mathscr{P}$ are definitely lost, as they will not be reconsidered again. Hence, the value assigned to $n^{min}_{\mathscr{P}}$ has to be carefully evaluated: Limiting the number of neurons in the pool will greatly reduce the computational cost of the rest of the algorithm, however, the risk of not including neurons that are actually part of a community increases. Throughout this paper we adopt the "safe" choice $n^{min}_{\mathscr{P}} = 1$, for which $\mathscr{P}$ coincides with the whole network when N and Z_B are sufficiently large like the ones we use. This is also the case of the toy network we are considering in this section (all neurons of the network are admitted to the pool, see Fig. 4.1A).

Neurons in $\mathscr{P}$ can be sorted in a *popularity ranking* based solely on $n^i_{\mathscr{P}}$ (Fig.4.1A, *right*). In doing so, nodes with the same value of $n^i_{\mathscr{P}}$ are treated as identical because we are (temporarily) losing all the detailed information of which pairs of neurons are effectively connected with each other. *Step 2* below is built only upon the popularity ranking, therefore without the need to access this detailed information. This allows to save a considerable amount of computational resources and to speed up the research, while still obtaining great results in terms of community detection.

Popularity ranking is a preliminary step, deterministic and with no approximations (i.e. there is no loss of information) as long as the threshold $n^{min}_{\mathscr{P}}$ is kept to a low value. From now on we will be working only with the neurons in $\mathscr{P}$.

Step 2. Blob search Our heuristic approach consists in using the popularity ranking to narrow the research to the regions in $\mathscr{P}$ where it is more likely to find a community. The key observation is that once we give a formal definition of community like Eq. (4.5), we can use it a basis for a reliable heuristic search. On the left hand side of Eq. (4.5) we have the number of bidirectional pairs formed by neuron i within the community $\mathscr{C}$, which, by using the notation introduced in this section, can be rewritten as $n^i_{\mathscr{C}}$. As pointed out earlier, these quantities are unknown; however, they are upper bounded by $n^i_{\mathscr{P}}$, which corresponds to the case where all the bidirectional pairs that a neurons forms are part of the same, unique community. We can revert this argument and derive from Eq. (4.5) the size of the largest community that each neuron can potentially form within the pool:

$$|\mathscr{C}^{max}_i| = \frac{n^i_{\mathscr{P}}}{\vartheta_{\mathscr{C}}} + 1 \qquad \forall i \in \mathscr{P}\,. \tag{4.6}$$

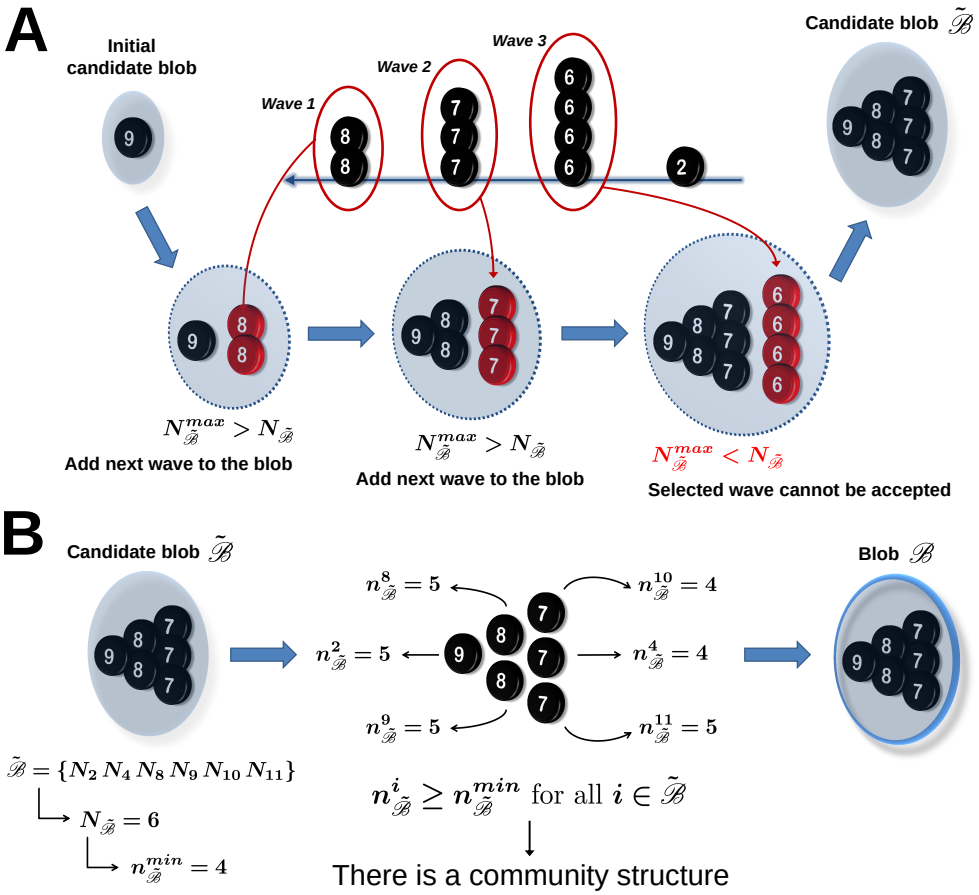

Figure 4.2 Algorithm Step 2: Blob search. A *Part 1*: Detecting a candidate blob $\tilde{\mathscr{B}}$. The research starts with the highest ranked neuron as the only one in $\tilde{\mathscr{B}}$ (*top left*). The other neurons left in the ranking are divided into waves depending on $n^i_{\mathscr{P}}$ and are progressively added to the candidate blob until the condition $N^{max}_{\tilde{\mathscr{B}}} > N_{\tilde{\mathscr{B}}}$ is violated (*middle*), see explanation in the text. If $N^{max}_{\tilde{\mathscr{B}}} < N_{\tilde{\mathscr{B}}}$, as in this example, then $\tilde{\mathscr{B}}$ is the set of neurons found before adding the current wave (*top right*). **B** *Part 2*: Candidate blob validation. The full community definition is restored within $\tilde{\mathscr{B}}$, giving the minimum number of bidirectional pairs that each neuron of the blob needs to form within the blob itself, $n^{min}_{\tilde{\mathscr{B}}}$ (*left*). In this example all the neurons in the candidate blob meet this requirement (*middle*), as they are all-to-all bidirectionally connected except for the pair N_4, N_{10}. The candidate blob satisfies then the complete community definition and gets the status of blob $\mathscr{B}$ (*right*).

$|\mathscr{C}^{max}_i|$ does not take into account which are the neurons connected with neuron i through the $n^i_{\mathscr{P}}$ bidirectional connections. As a consequence, all the $N_{\mathscr{P}} = |\mathscr{P}|$ relations of Eq. (4.6) are uncoupled, differently from what happens in the community definition. Starting from Eq. (4.6), this second step aims at filling the gap with Eq. (4.5) by progressively incorporating these element that have been discarded, hence producing as a result sets of neurons in $\mathscr{P}$ that satisfy the full community definition. The goal of this step is indeed to find the largest

possible communities in the network. This is done through a recursive two-step procedure, depicted in (Fig. 4.2):

- *Part 1.* Starting form Eq. (4.6), we find the largest set of neurons in the pool that could potentially form a community based only on the information contained in the popularity ranking. We call this a *bidirectional candidate blob* $\tilde{\mathscr{B}}$. Formally, $\tilde{\mathscr{B}}$ is a set of neurons such that:

$$n^i_{\mathscr{P}} \geq \vartheta_{\mathscr{C}} \left(|\tilde{\mathscr{B}}| - 1 \right) \quad \forall i \in \tilde{\mathscr{B}} \qquad \textit{with } \tilde{\mathscr{B}} \textit{ as large as possible.} \tag{4.7}$$

  Respect to Eq. (4.6), with Eq. (4.7) we are restoring the coupling between the equations, that is an essential feature of the community definition. The approximation that we are making respect to Eq. (4.5) is clear when we compare the two relations: Neurons are included in $\tilde{\mathscr{B}}$ not because of the number of bidirectional connections that they form within $\tilde{\mathscr{B}}$, as the community definition would require, but depending on the total number of bidirectional connections that they form in the entire pool, $n^i_{\mathscr{P}}$.

  Note that $\tilde{\mathscr{B}}$ does not coincide with the potential community that the most popular neuron can form ($\max_{i\in\mathscr{P}} |\mathscr{C}^{max}_i|$), but it is highly likely that such a neuron is part of $\tilde{\mathscr{B}}$. In other words, the most popular neuron $i^* = \text{argmax}_{i\in\mathscr{P}} |\mathscr{C}^{max}_i|$ is the node that has the highest probability in the entire network to belong to $\tilde{\mathscr{B}}$, hence it is the first one to be recruited. In the example, $i^* = N_2$, with 9 bidirectional pairs formed in the pool (see Fig. 4.2A, *top left*). The other neurons are organised in waves, formed by identically ranked nodes, that are evaluated one at a time in a descending order (Fig. 4.2A, *middle*). At every iteration, the candidate blob is fully characterised by two quantities: The actual size $N_{\tilde{\mathscr{B}}}$ and the size $N^{max}_{\tilde{\mathscr{B}}}$ of the largest possible community the entire set $\tilde{\mathscr{B}}$ can be part of, based on the popularity ranking. Since for each neuron this is given by Eq. (4.6), then, for the candidate blob as a whole, $N^{max}_{\tilde{\mathscr{B}}}$ is determined by the last wave of neurons included:

$$N^{max}_{\tilde{\mathscr{B}}} = \min_{i\in\tilde{\mathscr{B}}} |\mathscr{C}^{max}_i| \,. \tag{4.8}$$

  At the beginning, $\tilde{\mathscr{B}} = \{i^*\}$, hence $N_{\tilde{\mathscr{B}}} = 1$ and $N^{max}_{\tilde{\mathscr{B}}} = |\mathscr{C}^{max}_{i^*}|$. As we progressively recruit waves of neurons, $N_{\tilde{\mathscr{B}}}$ increases whereas $N^{max}_{\tilde{\mathscr{B}}}$ decreases. As long as $N_{\tilde{\mathscr{B}}} < N^{max}_{\tilde{\mathscr{B}}}$ then $\tilde{\mathscr{B}}$ can potentially be a community and we can keep on recruiting the next wave of neurons to investigate whether a larger candidate blob (which could lead to a larger community) is possible. When the inequality is no longer satisfied then the process of recruiting neurons stops. We can have two scenarios: $N_{\tilde{\mathscr{B}}} > N^{max}_{\tilde{\mathscr{B}}}$ means that some

neurons in $\tilde{\mathscr{B}}$, the most recently added ones, do not have enough bidirectional pairs: Even in the most optimistic case where all these pairs are within $\tilde{\mathscr{B}}$, these neurons do not meet the community threshold, given the actual size of $\tilde{\mathscr{B}}$. Thus, the largest possible candidate blob is the one found at the previous iteration. This is the case of the toy network we are considering, as shown in Fig. 4.2A, *top right*. The second scenario is when $N_{\tilde{\mathscr{B}}} = N_{\tilde{\mathscr{B}}}^{max}$, and the largest possible candidate blob is the set of neurons found at the current iteration.

- *Part 2.* At this stage we have a set of neurons, a candidate blob $\tilde{\mathscr{B}}$ (Fig. 4.2B, *left*), that satisfies Eq. (4.7): According to the number of bidirectional connections that each of them forms in the pool, $\tilde{\mathscr{B}}$ is suitable to form a community. We can therefore move on and restore the full community definition, Eq. (4.5), by computing the number of bidirectional connections $n^i_{\tilde{\mathscr{B}}}$ that each neuron in $\tilde{\mathscr{B}}$ forms within $\tilde{\mathscr{B}}$ itself (Fig. 4.2B, *middle*). In other words, we are substituting $n^i_{\mathscr{P}}$ with $n^i_{\tilde{\mathscr{B}}}$ in Eq. (4.7), obtaining exacly the community definition. Thus, if each neuron in $\tilde{\mathscr{B}}$ satisfies the condition, then there is a community structure and we call it a *bidirectional blob* $\mathscr{B}$ (Fig. 4.2B, *right*): A set of neurons that certainly contains at least one bidirectional community.

  If the community definition is not satisfied, then we withdraw those neurons that violate it to obtain a new $\tilde{\mathscr{B}}$ and to apply *Part 2* again. This refinement process continues until the algorithm finds a blob $\mathscr{B}$ or until $\tilde{\mathscr{B}}$ contains only one neuron, meaning that there is no blob.

Whenever this step gives a non empty blob as a result, then we proceed with *Step 3* below to finally find the communities that are present in $\mathscr{B}$. After this, we temporarily eliminate from the pool all the neurons that have been detected as being part of a community so far, and to this modified pool we apply *Step 2* from the beginning. Therefore, if a neuron is found to be part of a community, it does not get the chance to be evaluated again for being included in other blobs, meaning that blobs are all disjointed sets. This is one of the reasons why we introduce *Step 3* below, which is built to detect overlapping communities.

The procedure continues until there is no blob found. In this case the entire algorithm goes to an end and its final outcome are all the bidirectional communities found so far within the previously detected blobs. Thus, the result of this step is a set of non-overlapping blobs, each of them containing for sure at least one bidirectional community.

Step 3. Looking into a blob: The friendship algorithm The set $\mathscr{B}$ that we obtain at the end of *Step 2* certainly forms a community, as it satisfies the definition. However: *i)* Two different communities, or at least parts of them, may be detected within the same blob

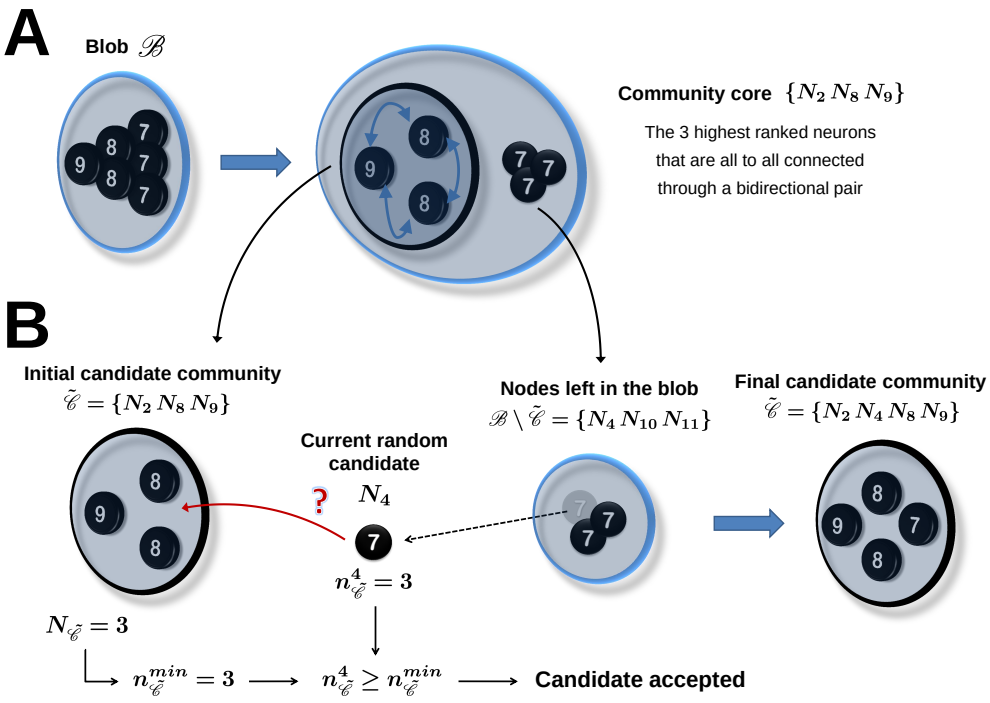

Figure 4.3 Algorithm Step 3: Friendship algorithm. A Detecting the candidate community core: The highest ranked neurons in the blob that are also all-to-all bidirectionally connected. **B** Building a candidate community: Starting from the core (*left*), one of the neurons that are left in the blob is randomly selected at a time and, based on the community definition, its inclusion in the candidate community is evaluated (*middle*). Here we show only the first iteration with the neuron N_4: Because it forms a bidirectional connection with all the 3 neurons in $\tilde{\mathscr{C}}$ ($n^4_{\tilde{\mathscr{C}}} = 3$), it can be accepted, resulting in a bigger candidate community for the next iteration (*right*).

(*resolution problem*). *ii)* The method does not correctly identifies overlapping communities. If there is an overlap between any of them, the above procedure assigns the intersection region to only one community (*overlapping problem*). *iii)* Mistakes may occur during the identification process: Some neurons that are originally part of a community may have been left out of the blob that contains that community, we call them *good friends*, whereas some other neurons may have been erroneously included in $\mathscr{B}$, we call them *false friends* (*accuracy problem*).

In order to address these issues, at this stage we define a procedure that builds up a community neuron by neuron through direct verification of the definition. This approach is much more feasible and robust at this stage rather than at the very beginning, for two reasons: *i)* We apply it to a very limited set of neurons, i.e. the blob $\mathscr{B}$. *ii)* We already know that these neurons are relevant in terms of community structure.

Step 3 starts by selecting the three neurons of the blob that are the most popular ones and at the same time form only bidirectional pairs among them. We call this *candidate community core*, and its purpose is to give a likely basis where to start building the *candidate community* $\tilde{\mathscr{C}}$ (Fig. 4.3A). The reason why we choose three neurons is to minimise mistakes: There is no guaranty that these three nodes really belong to the same community, but the probability of this event to happen is high, and it monotonically decreases with the size of the candidate community core. Also, among all the possible triplets of neurons, these three nodes maximise the probability of having three neurons being part of the same community, since they are the highest in the ranking. On the other side, the choice of having a core of only two neurons does not seem very reliable because, due to the random generation process of connections in the network (see below), the event that two neurons are connected with a bidirectional pair is very common, even among neurons that are not part of any community. Once we identify the candidate community core, at each iteration we randomly select a neuron in the blob and we ask if it satisfies the community definition (Fig. 4.3B): Given the current size of the candidate community $N_{\tilde{\mathscr{C}}}$, by using the definition Eq. (4.5) and rounding up the result, we have the minimum number of bidirectional connections $n^{min}_{\tilde{\mathscr{C}}}$ that the *candidate neuron* needs to form with the current members of $\tilde{\mathscr{C}}$ in order to join it.

In Fig. 4.3B we show the procedure for the first iteration only, where $\tilde{\mathscr{C}} = \{N_2 N_8 N_9\}$, hence $N_{\tilde{\mathscr{C}}} = 3$, $n^{min}_{\tilde{\mathscr{C}}} = 3$, and N_4 is the randomly selected node. This neuron forms 7 bidirectional connections in the entire network, but at this stage this is not relevant anymore. What matters is that it forms a bidirectional pair with each of the neurons in the current candidate community ($n^4_{\tilde{\mathscr{C}}} = 3$), meaning that it is "friend" with all of them and thus it can clearly be accepted in $\tilde{\mathscr{C}}$. In the next iteration, the candidate community is then $\tilde{\mathscr{C}} = \{N_2 N_4 N_8 N_9\}$, which results in $N_{\tilde{\mathscr{C}}} = 4$ and $n^{min}_{\tilde{\mathscr{C}}} = 3$ again. Thus, the neuron that will be selected, either N_{10} or N_{11}, needs to form at least 3 bidirectional connections with the 4 candidate community's members in order to join it. This is exactly what happens in this example, and, since in the last iteration also the last neuron turns out to have enough friends in $\tilde{\mathscr{C}}$ (for which it will be $N_{\tilde{\mathscr{C}}} = 5$ and $n^{min}_{\tilde{\mathscr{C}}} = 4$), the final candidate community will coincide with the blob: $\tilde{\mathscr{C}} = \{N_2 N_4 N_8 N_9 N_{10} N_{11}\}$.

Note that in early iterations, when the candidate community is not well formed yet, false friends still get a chance to pass the test and be recruited in $\tilde{\mathscr{C}}$. Conversely, strongly connected neurons within the original community have much higher chances to pass the test, no matter at which iteration they are selected. Thus, once the recruitment within the blob has finished and $\tilde{\mathscr{C}}$ is formed, we check again that each neuron is entitled to stay in $\tilde{\mathscr{C}}$ through the verification of the community definition (*false friends expulsion*). In the example we are

using, no neuron in $\tilde{\mathscr{C}}$ fails to meet the requirement so they all definitely earn the right of being in the candidate community.

Most likely, the above procedure returns a set that is a single, well consistent candidate community, hence solving the resolution problem. However, it might not be the original complete community: Good friends may be left out, most likely in the blob $\mathscr{B}$ but also in the pool $\mathscr{P}$. Therefore, we start recruiting for a second time with the condition of satisfying the community definition, first among the neurons left in the blob and then within the pool (*good friends inclusion*). This procedure, together with the false friends expulsion, address the accuracy problem. Moreover, recruitment among neurons in the pool allows to consider for the current candidate community also neurons that have been previously included in other blobs or communities, hence solving the overlapping problem. In our example, there are no neurons left in the blob so we check the ones in the pool: N_1, N_3 and N_7 are recruited whereas N_5 and N_6 are left out, giving finally a community candidate of 9 neurons that coincides with the original true community.

Step 4. Candidate community global control As pointed out earlier, based on Esposito et al. [2014], for a set of neurons to be a bidirectional community $\mathscr{C}$, its symmetry measure Eq. (4.1) must exceed a threshold value s_B, which depends on the distribution of connections considered. Now that we isolated a candidate community from the rest of the network, we are in the position of applying this criterion. Candidate communities that do not pass this test are sets that cannot be qualified as communities. Note, however, that they are still bidirectional communities in the sense of our topological definition Eq. (4.5). Since this definition is threshold-based, it introduces a binary criterion with subsequent loss of information. The definition is a guidance for community detection that reduces the weighted network to an un-weighted one. Thus, once the research has been successful, the complete information stored in the weights needs to be recovered and the actual identity as bidirectional community can be finally evaluated by means of the symmetry measure.

Sets that cannot be qualified as communities present an excess of bidirectional pairs due to the random generating process, which made these sets to be detected as possible communities, but failure in the symmetry measure test means that the rest of the pairs are far from being bidirectional, hence pulling the value of the symmetry measure down within the randomness boundaries. This is no evidence that learning took place in the specific set as a whole. Statistically, this situation is likely to happen for small sets of neurons, and indeed this is when we observe failure of the symmetry measure test. These sets of neurons are therefore safely withdrawn.

Step 5. Noisy candidate communities identification Besides the real communities that are present in the network as a result of learning, communities can be also formed out of chance, due to the randomness in the network's connectivity: It is highly likely that small sets of $4-5$ neurons show community properties and thus will be detected as such. Since the probability of randomly forming a community dramatically drops with the size, we can define a *noise threshold* ϑ_{noise} and discard all sets below such a threshold. This clearly fixes a lower limit to the resolution of the algorithm. However, the maximum size of a random community, which ideally corresponds to such a threshold, grows with the size of the network in a way that is sub-linear, allowing to set a unique, relatively small threshold for all the networks with a large size that does not affect the overall performance.

Step 6. Single community reduction At this stage we have single communities $\mathscr{C}$, but we might have a final *redundancy problem*, especially for networks of small size: It can happen that the same original community has been detected more than once, every time with different false friends included and good friends left out. It can also happen that the original community is broken into two overlapping parts detected as different communities. The final step is therefore trying to resolve these issues calculating the overlap degrees between each pair of communities. Pairs with an overlap exceeding the *overlap threshold* ϑ_{ω} are merged together and the symmetry measure is used as an evaluating criterion: If the value on the merged community is higher than the values of the single communities, then the merged community definitely replaces the two single ones, otherwise they are kept separate.

4.1.4 Network and communities generation: Benchmark procedure

To test the above algorithm, we generated *in silico* data representing several different scenarios, which will be discussed in the Results section alongside the algorithm performance. Below, we describe the general procedure used to produce a connectivity matrix for a network containing bidirectional communities.

The first step for creating a bidirectional community is to decide the value of its symmetry measure, which has to be in the range $[s_B, 1]$ (Esposito et al. [2014]). By definition, this gives the mean value of the strength of the connection pairs in the community, $\langle Z \rangle = 1 - s$. Because the learning process shapes the connections of a community in the same direction, it is reasonable to assume that at the end of learning the pairs form a Gaussian distribution. Therefore, for each community in the network, we generated the set of $\{Z_{ij}\}$ according to a Gaussian distribution with mean $\langle Z \rangle$ and standard deviation σ, that is a free parameter. We recall that Z is a variable ranging from 0 to 1 and that the bidirectionality region is $[0, Z_B]$. Based on this, two issues may arise when we generate the pairs, related to the two boundaries

and to the choice of σ: *i)* Some of the Z_{ij} could be negative. If this is the case, the tails of the distribution are symmetrically folded towards the inside so as to guarantee the non-negativity of the $\{Z_{ij}\}$ and to preserve the mean value of the distribution. *ii)* A considerable part of the distribution could fall in the randomness domain ($Z > Z_B$), meaning that many pairs will not be classified as bidirectional. As a consequence, some neurons may not form the minimum number of bidirectional pairs required by the community definition, resulting in the whole set not being a community anymore. To avoid this issue, we make sure that the integral of the Gaussian in the bidirectional region, which gives the probability of forming a bidirectional pair $\wp_B$, exceeds the community threshold $\vartheta_{\mathscr{C}}$.

Once we have the set of $\{Z_{ij}\}$ for each community, we can generate the single connections w_{ij}. A first half of them is directly drawn from the uniform distribution in $[0,1]$, be the upper (or the lower) triangular part of the community's connectivity matrix. This first half, together with the $\{Z_{ij}\}$, is used to compute the second half of the single connections by means of Eq. (4.2). The rest of the connections in the network are drawn from the uniform distribution in $[0,1]$.

Overlaps between communities are governed by the set of parameters $\{\omega_{\lambda\rho}\}$ representing the fraction of the community ρ that is in common with the community λ:

$$\omega_{\lambda\rho} = \frac{|\mathscr{C}_\lambda \cap \mathscr{C}_\rho|}{|\mathscr{C}_\rho|} \tag{4.9}$$

We allow overlaps only between subsequent pairs of communities. In other words, we can progressively enumerate the communities in the network in such a way each of them overlaps at most with only the previous and following one. Formally: $\omega_{\lambda\rho} = 0$ if $|\lambda - \rho| \geq 2$, leading to a tridiagonal matrix of overlaps. In cases of overlap between two communities, after having generated the first community, the mean of the pairs in the intersection is computed and it is used to offset the mean of the Gaussian distribution for the rest of the pairs in the second community, so as to preserve the value of the symmetry measure that we chose.

The set of parameters $\{s_\lambda\}$, $\{\sigma_\lambda\}$, $\{\omega_{\lambda\rho}\}$ we introduced here for the connections generation, together with the size of communities $\{N_\lambda\}$ and network N, entirely define the structure of a network, but they do not uniquely determine its connectivity because all the connections are generated through the above mentioned random process. Due to the presence of random elements in both data generation and detection procedure, for each combination of parameters we consider, we repeat the experiment $n_{iter} = 100$ times. Each experiment, or run, consists in generating the network connectivity as described above and applying our detection algorithm. Cumulative and averaged results are displayed in the appropriate section.

4.1.5 Analysis of the results: Measuring successful detection

At the end of each run, on one side we have the communities that we generated at the beginning, i.e. the real communities, and on the other side those detected by the algorithm. One way of measuring the quality of the results is to count how many good neurons have been detected. However, since this is going to be displayed as an average over the runs, we may lose too much information about the single run. Also, as pointed out earlier, failure in detecting single neurons may still happen despite the bulk of the community has been correctly identified.

Therefore, we introduce a criterion to determine successful community detection: Whenever the number of neurons in a detected community equals at least a fraction ϑ_{recog} of the neurons in a real community, we count that real community as successfully identified. If there is more than one detected community for which this happens relatively to the same real community, then the one with the highest percentage is considered to be one matching the real community and the others are counted as false communities, unless they result to match some other real community in the network. We choose $\vartheta_{recog} = 75\%$ as in our opinion three quarters is a fraction that already carries the distinctive features of the community to which it belongs.

At the end of the results' evaluation, the analysis of the algorithm performance can be done by using he following information on each real community: *i)* How many times it has been successfully detected in all runs. *ii)* How many good neurons have been identified as average across the runs. Alongside, we also display information about false communities that have been detected and false neurons included in good communities. Results about communities detection provide a quantitative tool to evaluate the goodness of the algorithm, whereas neurons detection provide a qualitatively information on its accuracy. Finally, we show the time needed to run the algorithm.

4.1.6 Thresholds

The algorithm described above makes use of 5 customisable thresholds, see Tab. 4.1. Throughout this paper we keep them fixed at their respective values. Since we assume a uniform distribution of connections prior to learning, we can directly rely on the results of Esposito et al. [2014] for uniform distributions: By fixing a level of confidence at $p = 0.05$, the bidirectionality threshold we use is $s_B = 0.6954$, which in turn gives $Z_B = 0.3046$. The threshold $\vartheta_{\mathscr{C}}$ for community existence is rather arbitrary and it can be fixed according to how dense we require the communities to be. In the present study we choose $\vartheta_{\mathscr{C}} = 75\%$. Concerning the noise effect, after observing the size of the noisy communities detected by

the algorithm, we fix $\vartheta_{noise} = 30$. For the other thresholds, also arbitrary, we use $n_{\mathscr{P}}^{min} = 1$ as a safe choice (as previously stated) and $\vartheta_{\omega} = 25\%$ as a limit case before two communities can be considered as part of a single bigger one (after evaluation of symmetry measure, see *Step 6*).

4.2 Results

In this section we present the results obtained by applying the community detection algorithm to networks of neurons with different community structures. In all the cases, we assume that the given network is the final product of a learning process that shaped the connections of some sub-regions away from the initial uniform distribution, to form what we called bidirectional communities, Eq. (4.5). The rest of the connections remain unchanged and therefore they are uniformly distributed. Network connectivity is generated according to the procedure outlined in Methods section.

Since the learning process is not explicitly simulated here, we have total control on the final structure of the network, through the tuning of 6 sets of parameters: The size of the network N, the number of communities ν, the size of communities $\{N_\lambda\}$ and the overlap between communities $\{\omega_{\lambda\rho}\}$ define the architecture of a network. The symmetry measure of the communities $\{s_\lambda\}$ and the standard deviation of the connection pairs in the communities $\{\sigma_\lambda\}$ define how much a community has been shaped towards bidirectionality.

The way the algorithm is constructed, we expect that strong bidirectional communities, i.e. with $s \to 1$ and small σ, are the easiest to detect, compared with bidirectional communities with $s \to s_B$ and large standard deviation. The degree of difficulty in the detection of a community can be derived from the way we generate the community itself. Indeed, as pointed out in the Methods section, s and σ determine the number of bidirectional pairs that each neuron forms in the community, through a random process. Thus, necessary condition for a set of neurons to be a community (and therefore to be detected) is that this number exceeds the threshold for community existence. Also, the closer this number to the threshold (from above), the more difficult the detection of the community.

Each simulation consists of $n_{iter} = 100$ runs, each of them starts with communities and network generation, continues with the communities detection algorithm and finally ends with the evaluation of the results, where we compare detected and generated communities. To evaluate the algorithm performance, we use two indicators for each community generated in the network. The first quantity counts how many times a community has been successfully detected during the n_{iter} iterations (see Methods). Once a community has been correctly identified, the second indicator measures how many nodes of the generated community have

been detected, and displays this information as an average percentage across n_{iter} iterations and relative to the total number of nodes in the generated community. Additional indicators for the number of false communities detected and for the percentage of false neurons in a correctly detected community complete the evaluation.

4.2.1 Networks with a single community

Alongside the size of the community, we introduce the community to network ratio $r_{c/n} = N_{\mathscr{C}}/N$, which is a more significant indicator to assess the algorithm performance. A complete evaluation (at least in the single community case) requires, therefore, carrying out 3 different analyses, corresponding to fixing one of the three quantities $N_{\mathscr{C}}, N, r_{c/n}$ while varying the other two.

Single, size-varying community in a network of fixed size

We begin the analysis of the algorithm performance with the most common and typical problem: Given a network, we want to know if there is a community and of which size. Therefore, we fix the size of the network at $N = 5000$ neurons and we systematically vary the size of the community in the set $\{75, 100, 250, 500, 750, 1000, 2500\}$. The community is generated with $s_{\mathscr{C}} = 0.75$ and $\sigma_{\mathscr{C}} = 0.05$, resulting in a probability of forming a bidirectional pair $\wp_B = 0.863$. Correctly, $\wp_B > \vartheta_{\mathscr{C}}$ (see Network and community generation subsection in Methods).

In Fig. 4.4A we report the results concerning community detection: Blue bars show the cumulative number of successful detection of the generated community, whereas the upside down red bars count the number of false communities. Similarly, Fig. 4.4B shows the average percentage of good neurons (blue bars) and false neurons (upside down red bars) in the detected community. Blue bars carry what we can call a *positive information* as we want to maximise them, whereas red bars is what we want to minimise to zero, hence they carry a *negative information*. The horizontal black dashed line marks the optimality level for positive information: When for the same value of $r_{c/n}$ both bars of Fig. 4.4A,B hit this level means that the algorithm has detected all the neurons forming that community all the time. For the values considered here, this is almost always the case, except for the smallest community case where detection of the community is successful only 20% of the times. Note that whenever this community is identified, the algorithm correctly recruits all the good neurons (the blue leftmost bars of Fig. 4.4A,B have the same height). As expected, these results suggest that the bigger the size of the community the easier to detect it, with a critical value of ~ 100 neurons.

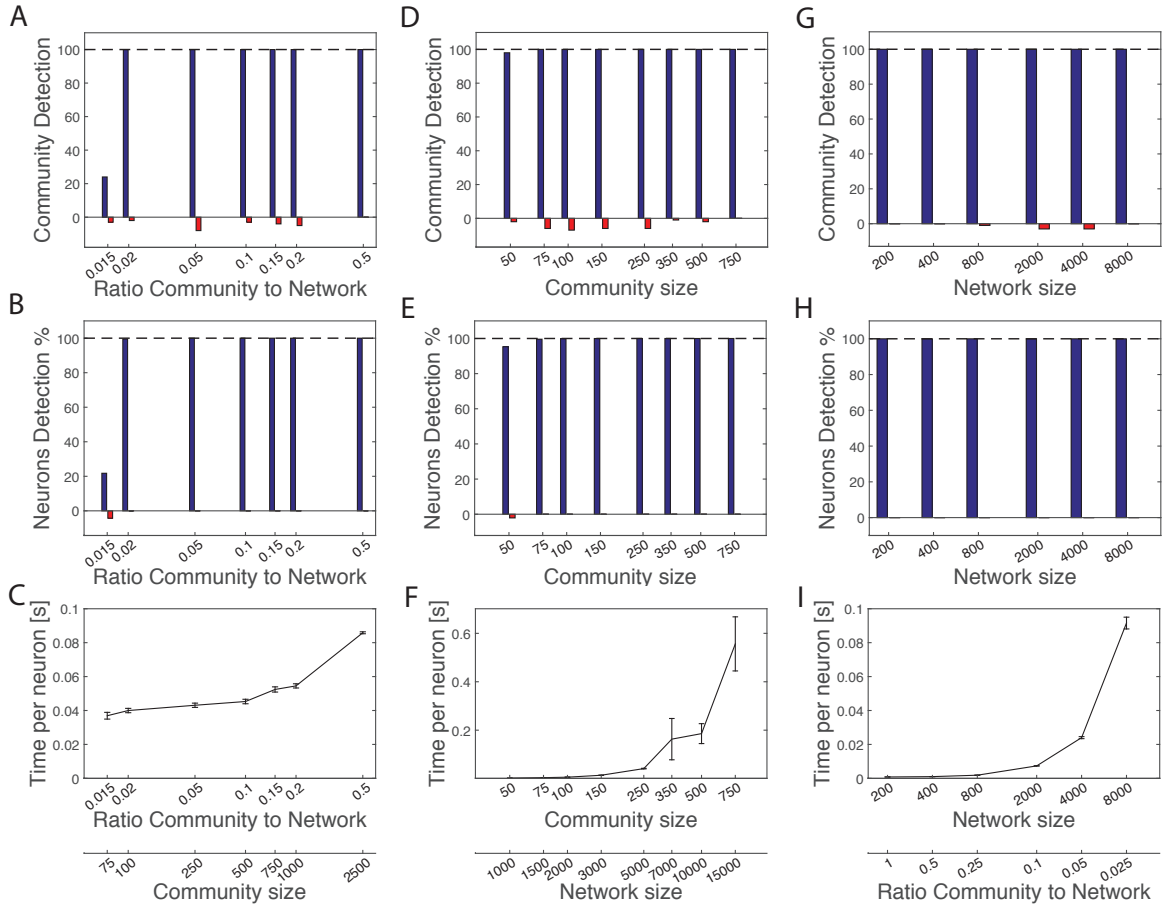

Figure 4.4 Algorithm Performance for the case of single community of $s_{\mathscr{C}} = 0.75$ and $\sigma_{\mathscr{C}} = 0.05$ embedded in a network. **A - C** Network' size fixed at $N = 5000$ neurons while varying community' size. **D - F** Ratio community to network fixed at $r_{c/n} = 0.05$ while varying both network' and community' size. **G - I** Community' size fixed at $N_{\mathscr{C}} = 200$ neurons while varying network' size. All results in each panel are relative to $n_{iter} = 100$ repetitions. **A, D, G** Cumulative community detection. *Blue bars:* Successful detection, *Red upside down bars:* False detection. The dashed line represents the best possible performance of correctly detecting the community all the time. **B, E, H** Average percentage of neurons detection, relative to the size of the generated community. *Blue bars:* Good neurons, *Red upside down bars:* False neurons. **C, F, I** Simulation time per neuron in the network. Error bars represent standard error. Note that the scale of all x-axes is logarithmic.

Fig. 4.4C shows a very interesting result, direct consequence of the algorithm architecture: The computational time per neuron in the network is smaller for small communities. In other words, when we increase the difficulty of the task, the time needed for the detection is reduced, provided that the size is above the critical value for the search to be successful. It is also interesting to note that if we increase the bidirectionality of the community (by increasing the value of its symmetry measure), the algorithm time is the same (result not

shown here). This means that detection time is not affected by the internal structure of the community but only by its size.

Single, size-varying community in a size-varying network with fixed ratio

The above scenario gives only partial information on the goodness of the algorithm, as the size of the network is fixed to a single value: Fig. 4.4A-C show results when we vary only the size of the community to account for different ratios community to network. However, the task of finding a community of 100 neurons in a network of 1000 of nodes might be different from searching for a community of 1000 in a network of 10000 nodes.

Therefore, we investigate a second scenario: A single, size-varying community within a network whose size also varies, in such a way to keep the ratio community to network fixed. We choose a relatively small value $r_{c/n} = 0.05$, while the size of the community varies between the values $\{50, 75, 100, 150, 250, 350, 500, 750\}$. The size of the network varies accordingly from 1000 to 15000. As above, $s_{\mathscr{C}} = 0.75$ and $\sigma_{\mathscr{C}} = 0.05$.

Results are shown in Fig. 4.4D-F, with the same meaning of quantities and colours as in Fig. 4.4A-C. Detection is perfect almost all the time, with very few mistakes mostly in the sense of detecting false communities. At first, the performance is fairly independent of the absolute sizes, as expected. A more careful inspection shows that slightly better performances are obtained for larger sizes. The reason could be that for larger networks the fluctuations on the bidirectional pairs formed out of chance become smaller and for the neurons being part of the generated community is easier to stand out of the crowd of nodes, hence the precision of the algorithm increases.

Single, fixed size community in a size-varying network

Finally, to complete the analysis of the single community case, we study the algorithm performance when we increase the size of the network while keeping fixed the number of nodes in the community. We choose a small community of $N_{\mathscr{C}} = 200$ neurons and we vary the size of the network in the set $\{200, 400, 800, 2000, 4000, 8000\}$. The ratio community to network varies accordingly from 1 to 0.025. Again, $s_{\mathscr{C}} = 0.75$ and $\sigma_{\mathscr{C}} = 0.05$.

Results are shown in Fig. 4.4G-I. Once again, the performance of the algorithm is excellent in the range of values considered, in terms of both positive and negative information. In particular, Fig. 4.4H shows that the algorithm finds exactly the 200 neurons forming the community all the time, with no false neurons. As expected, increasing the size of the network also increases the time needed for the detection, with a dependence from the time per neuron of the network that looks quadratic.

Community	$N_{\mathscr{C}}$	$s_{\mathscr{C}}$	$\sigma_{\mathscr{C}}$	$\omega_{\mathscr{C}\mathscr{C}'}$	$\wp_B$
1	200	0.75	0.05	-	0.86
2	200	0.75	0.05	0.2	0.86
3	500	0.74	0.05	0.1	0.81
4	150	0.74	0.05	0.2	0.81
5	150	0.79	0.1	0	0.81

Table 4.2 List of parameter's values used to generate the community structure in the case of $\nu = 5$ communities. *Column 1:* Community progressive number. *Column 2:* Size of the community. *Column 3:* Symmetry measure. *Column 4:* Standard deviation of the connection pairs Z. *Column 5:* Overlap with the previous community, expressed as number of common neurons divided by the number of total neurons in the community. *Column 6:* Probability that a neuron of the community forms a bidirectional pair, as a result of a Gaussian distribution with parameters based on Columns 3 and 4.

4.2.2 A multiple communities case

Finally, we wish to study the behaviour of our detection algorithm when more than one community is present in the same network. As an example, we choose a challenging task: A network with 5 communities generated with different parameters' values so as to have a certain complexity in the overall structure, see Tab. 4.2. The size of network is varied in the set $\{1500, 2143, 3000, 5000, 7500\}$. Note that values of the symmetry measure are all very close to the limit between bidirectionality and randomness, making the detection more difficult, as can be inferred from the last column of the table.

Fig. 4.5A shows the global performance of the algorithm, in the form of stacked bars for each value of the network's size considered. The *dark grey* part at the bottom of the bars counts how many times the 5 communities have been correctly detected as 5 different communities, as an average over $n_{iter} = 100$ runs. The central part in *grey* shows the average number of times that a community has been detected as an unresolved community, i.e. two overlapping communities detected as a single big one (never more than two). The upper part in *light grey* shows the average number of false communities. Clearly, as the size of network increases, the number of false communities also increases, but the performance on the 5 true communities remains stable and optimal. Indeed, all the communities are detected almost all the time, either resolved (more than 95% of the time) or unresolved. From a more detailed analysis of the results, it can be seen that when there is an unresolved community this is always the union of the communities 2 and 3 in Tab. 4.2. This is reasonable as $\mathscr{C}_3$ is by far the largest in the network and the overlap of 0.1 with community $\mathscr{C}_2$ means that they have 50 neurons in common. From the perspective of community 2 this is a considerable overlap of 25%, which indeed equals the value we chose for the overlap threshold ϑ_ω. It is therefore a

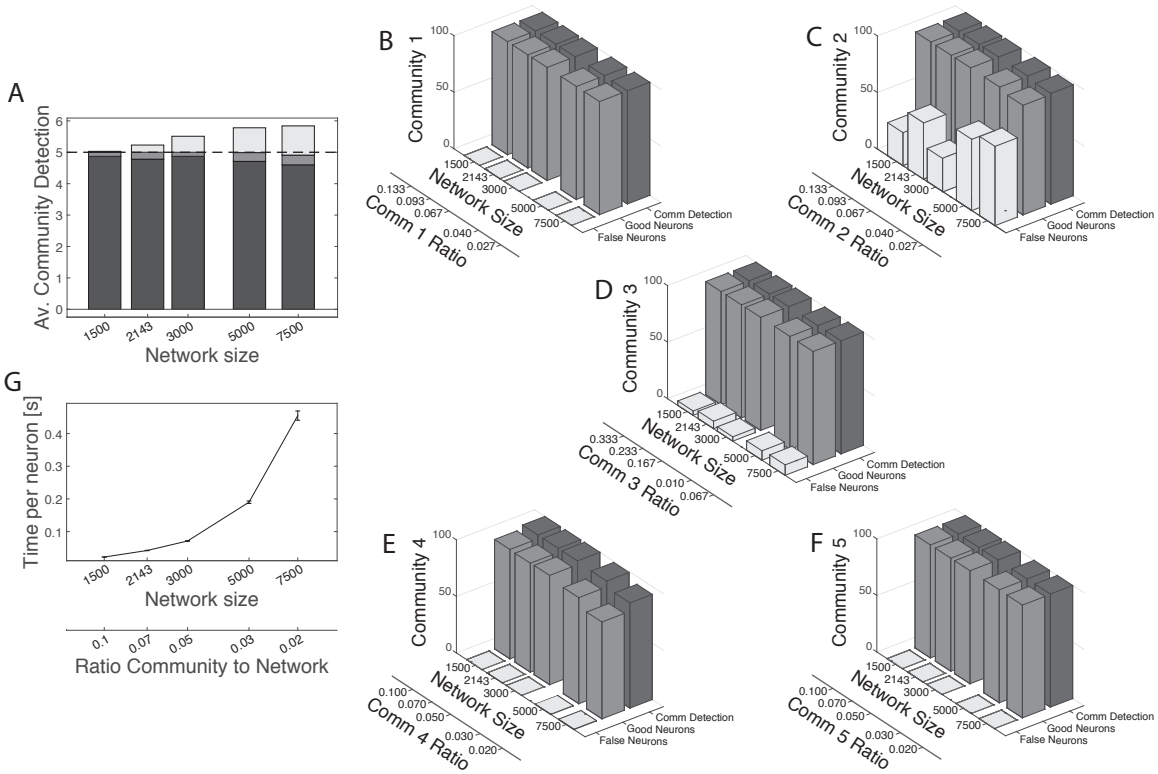

Figure 4.5 Algorithm Performance for a network with complex structure. Five communities with different sets of parameters, see Tab. 4.2, are embedded in the network. Communities' size are kept fixed while varying network' size. All results in each panel are relative to $n_{iter} = 100$ repetitions. **A** Global performance. *Dark grey bars:* Successful detection with all communities resolved, *grey bars:* Successful detection with two communities unresolved (see text for details), *light grey bars:* False communities. The dashed line represents the best possible performance of correctly detecting all the five communities all the time. **B - F** Single community detection statistics. *Dark grey bars:* Cumulative community detection over the 100 repetitions, *grey bars:* Good neurons, *light grey bars:* False neurons. Note that the (expected) discrete amount of false neurons detected in communities 2 and 3 is due to the unresolved cases between these two communities (see text for a discussion). **G** Simulation time per neuron in the network. Error bars represent standard error. Note that the scale of all x-axes is logarithmic. The ratio community to network below panel g is relative to the smallest community in the network.

matter of few neurons whether these two communities are merged or not during the last step of the algorithm (see Methods).

Fig. 4.5B-F shows the algorithm performance community by community, with bars showing the cumulative community detection (*dark grey*), percentage of good neurons (*grey*) and percentage of false neurons (*light grey*). Detection of communities 1 and 5 is perfect, both in terms of good neurons and false neurons. Communities 2 and 3 present a visible amount

of false neurons, most of which, however, are due to the unresolved cases between these two communities. In this sense they are not completely false neurons. Finally, community 4 is the only one showing a decrease in the performance as the size of the network increases, with percentages still above 85% for $N = 7500$. The reason is that $\mathscr{C}_4$ is the most difficult to detect because is the one with the lowest values of $N_{\mathscr{C}}$, $s_{\mathscr{C}}$ and $\sigma_{\mathscr{C}}$.

The last panel, Fig. 4.5G shows the simulation time per neuron in the network. Note that, compared to the case of a single community, the time needed is nearly five times larger, suggesting that it grows linearly with the number of communities.

4.3 Discussion

In this paper we address the problem of structure detection in networks of neurons, which is of crucial importance in the study of connectome in Neuroscience, framing it within the well-known, in Network Theory, community detection problems. By nature, networks of neurons are weighted and directed graphs, which makes the problem of structures searching in these networks one of the most difficult ones to approach. In the general, for most of the problems related to structures detection in Graph and Networks Theory it is not possible to give an exact solution, hence heuristic algorithms are often adopted.

Here we present an algorithm for the detection of a particular class of communities in large scale simulated networks. Thus, this is intended mainly as a tool to help *in silico* research aiming at understanding the connectome. In the future, the opportunity of having direct access to synaptic weights, and therefore to connectivity matrices, may also allow a direct application of the algorithm to experimental data. Moreover, the algorithm could be of more general interest for pure studies in Networks and Graph Theory and it can be adapted to similar problems in other disciplines where nodes are not neurons.

Differently from cliques, that are very well defined objects, the concept of community is vague and we cannot find a unique definition in the literature. Traditional methods for community searching are either based on cliques or define a community *a posteriori*, i.e. as the outcome given by the algorithm. Here we propose a different approach by giving a formal definition for communities *a priori*. We show how having such a definition is an advantage as the algorithm can use it directly for a more efficient and direct research. Also, the method is based on a definition of symmetry measure, which allows manipulating the original connectivity and deal with quantities carrying much reduced information. This implies a loss of information, but we show that results are excellent and there is a great benefit in terms of time and computational resources.

The algorithm we present is based on statistics and it requires that the distribution of the connections is known and is somehow regular: We assumed a uniform distribution, but in principle the method works for any kind of distribution for which it makes sense to define mean and variance. As such, differently from traditional approaches, our method works better for large number of neurons: We show that increasing both the size of the network and the community gives better results. Also, we chose to focus on bidirectional communities, but the procedure can be extended to other kind of communities, for instance unidirectional ones. Indeed, similarly to bidirectional structures, experimental results show also an excess of unidirectional connections in some parts of the brain (Lefort et al. [2009]; Pignatelli [2009]; Wang et al. [2006]). Generalisation to sparse networks should be also possible.

The results we present here are relative to worst case scenarios, because the communities are generated with symmetry measure very close to the random domain. For communities that are more markedly bidirectional the performance would be even better. Besides measuring the success on community detection, the performance of the algorithm is also evaluated on the false detections: False communities and false neurons within good communities. Based on initial results, to minimise false communities we naively fixed a threshold for a minimum community size at 30 neurons, considering everything smaller as an outcome of the random process used to generate the connections in the network. This limits the resolution of the algorithm: If there are real communities whose size is smaller than the threshold, they will not be detected. Since the average size of communities formed out of random depends on the size of the network, this part of the algorithm can be improved by using a threshold that is a function of N. Such a function can be derived by carrying out a systematic analysis on completely random networks, both theoretically and through simulations.

It is worth noting that there is nothing that makes a false community different from a true community, except for the fact that the latter has gone through a learning process. A possible approach to solve this problem could be therefore to track the evolution of the connectivity whenever possible. After having detected the communities, looking back at the history of their connections could give important insights about which sets really experienced a learning process.

The algorithm we presented requires setting a number of thresholds, which makes the research highly customisable and it also allows different degrees of searching: For instance we can be interested in finding only highly significant bidirectional communities, if any. We can tune the thresholds as we like for a stricter search, which would require also less computational time. Once we have the outcome, we can then gradually relax the values of the thresholds for a broader search, if we need.

The full algorithm is the combination of two sub-algorithms executed one after the other. The first sub-algorithm alone already gives excellent results, especially for single community detection, with a massive reduction of the running time. Indeed, the second sub-algorithm is essential for dealing with overlaps and for resolving two communities that have been detected as a big one. Also, due to the statistical approach of the algorithm, it is possible to evaluate the level of noise (bidirectional pairs formed out of chance) and take it into account from the beginning of the procedure. This would allow to withdraw a consistent fraction of neurons before executing the two sub-algorithms and therefore to greatly reduce the number of neurons for the search. These aspects need to be further investigated, together with the possibility of introducing parallel computing, to improve on the computational time requirements.

Summary The algorithm introduced here aims at identifying bidirectional communities within networks of neurons. This study complements the work presented in the previous chapter, and altogether these two papers furnish a standalone piece of research that, despite far from being exhaustive, addresses one aspect of the *how* problem formulated in the introduction of the thesis. In the last chapter, improvements and possible adaptations of the algorithm to different scenarios are discussed. Before this, in the next chapter, the attention is turned to the *why* question: What lays behind the observed correspondences between motifs and synaptic properties?

Chapter 5

Adaptation of Short-Term Plasticity Parameters via Error-Driven Learning May Explain the Correlation between Activity-Dependent Synaptic Properties, Connectivity Motifs and Target Specificity[1]

It is the current belief that experiences and memories are registered in long-term stable synaptic changes. Long-term plasticity, and in particular Hebbian learning or Spike-Timing-Dependent-Plasticity (STDP), is a form of unsupervised learning that captures correlations in the neuronal input. Hence, their involvement in, for instance, the development of receptive fields (Clopath et al. [2010]; Song et al. [2000]) or memory and associations is long-standing knowledge. However, the variety of different long-term plasticity rules (Markram et al. [2011]), indicates that the precise synaptic prescriptions of long-term plasticity mechanisms remain unclear.

On the contrary, short-term plasticity (STP) is well described (Costa et al. [2013]; Le Be' and Markram [2006]; Markram et al. [1998b]; Rinaldi et al. [2008]; Romani et al. [2013]; Testa-Silva et al. [2012]; Varela et al. [1997]) in the context of specific models (Hennig [2013]; Rotman and Klyachko [2013]; Tsodyks and Markram [1997]). Its role in neuronal computation is assumed to be related to temporal processing, see for instance Natschläger

[1] This chapter is almost entirely taken from Esposito et al. [2015]

et al. [2001] or the work by Carvalho and Buonomano [2011], where STP is demonstrated to enhance the discrimination ability of a single neuron (i.e. a tempotron, see Gütig and Sompolinsky [2006]), when presented with forward and reverse patterns. Synapses with short-term plasticity are also optimal estimators of presynaptic membrane potentials (Pfister et al. [2010]).

At the microcircuit level (Binzegger and Douglas [2004]; Douglas and Martin [2007a,b]; Grillner et al. [2005]; Silberberg et al. [2005]), the non-random features of cortical connectivity have recently raised a lot of interest (Perin et al. [2011]; Song et al. [2005]) and experimental evidence seems to reveal a correlation between specific motifs and synaptic properties (Buonomano and Merzenich [1998]; Perin et al. [2011]; Silberberg and Markram [2007]; Song et al. [2005]; Wang et al. [2006]). Specifically, observations of young ferret cortices (Wang et al. [2006]) indicate that neurons connected by synapses exhibiting short-term facilitation form predominantly bidirectional motifs; conversely, neurons connected by synapses exhibiting short-term depression form unidirectional motifs. Interestingly, the same overexpression of connectivity motifs has been observed in another brain area, i.e. the excitatory microcircuitry of the olfactory bulb (Pignatelli [2009]).

Earlier work by Vasilaki and Giugliano [2012, 2014] attempted to shed light on this correlation between short-term plasticity and the observed wiring diagram configuration. They demonstrate that all facilitating or all depressing networks, upon receiving the same wave-like stimulation, give rise to the experimentally observed motifs: Bidirectional for facilitating synapses and unidirectional for depressing synapses. This was explained both in the context of mean field analysis and microscopic simulations as a frequency-dependent effect. This is a simple consequence of the type of input (wave like) and the choice of the STDP triplet rule (Pfister and Gerstner [2006]). Differently from the classical pair rule, the triplet rule displays a frequency-dependent behaviour, which can explain some experimental results (Sjöström et al. [2001]): At low frequencies the rule reveals the classic STDP and, given a wave-like input, it results in unidirectional connectivity (Clopath et al. [2010]; Vasilaki and Giugliano [2014]). At high frequencies, however, it reveals "classic Hebb" behaviour: Neurons that fire together, wire together. Hence, the low firing network develops unidirectional connectivity, while the high firing network develops bidirectional connectivity; for details see Vasilaki and Giugliano [2014]. However, the observed synaptic development was not associated to any particular type of learning, but was explored as the emerging structure upon receiving a wave like input: What the network learned per se in that context was not clear.

With the present work we aim to complement and extend on Vasilaki and Giugliano [2012, 2014]. We define a learning model for STP through which a population of neurons

can modify its synapses in order to adapt its own activity and then fulfil a given time-varying task. The key idea comes from an optimisation perspective: Neurons that are able to modify their synapses, for instance making depressing synapses more and more depressing or even turning them into facilitating ones, would allow for much more flexibility and efficacy in signal transmission. A similar argument can be found in Markram et al. [1998a], whereas for earlier but different mechanisms of STP optimisation or learning we redirect to Carvalho and Buonomano [2011]; Natschläger et al. [2001].

Moreover, supporting our hypothesis, recordings form rats show that sensory experience can modify short-term dynamics of excitatory synapses in somatosensory cortex (Cheetham et al. [2007]; Finnerty et al. [1999]) and that a developmental switch from depression to facilitation occurs in neocortical circuits (Reyes and Sakmann [1999]). Even more significative, experimental results revealed that recovery from short-term depression is activity dependent (Fuhrmann et al. [2004]): a decrease in short-term depression following high rate stimulation has been observed in Purkinje cells (Dittman and Regehr [1998]) as well as in excitatory neurons (Stevens and Wesseling [1998]; Wang and Kaczmarek [1998]). The most accepted explanation for this effect resides in an elevated concentration of calcium ions in the presynaptic terminal (Fuhrmann et al. [2004]; Hosoi et al. [2007]; Wang and Kaczmarek [1998]). This, in turn, is thought to either enhance vesicles replenishment in the release pool (modelled with an increase in the occupancy of the release pool (Wang and Kaczmarek [1998])) or decrease the time constant of recovery from depression (Fuhrmann et al. [2004]). This latter mechanism is included in our model and provides evidence for modification of STP parameters. We stress, however, that the model we present here is highly speculative as, to the best of our knowledge, the available experimental evidence is very little. Therefore, rather than developing a detailed and biologically plausible mechanism, we want to provide a model that, inspired by biology, could be applied to Articificial Intelligence and Neural Networks.

We construct a typical inverted associative learning problem (Asaad et al. [1998]; Fusi et al. [2007]; Vasilaki et al. [2009b]) where neurons have to learn to respond with high or low frequencies, when presented with the same wave-like input signal. We use this paradigm to show the potential of our model. In particular, not only do we provide an explanation for the correspondence motifs-synaptic properties within the context of learning both STP and STDP (triplet rule) but we also qualitatively capture, for instance, the heterogeneity in synaptic properties observed by Wang et al. [2006].

Moreover, having defined the learning model as a target-specific mechanism, we are able to obtain variability in the short-term profile of synapses innervating functionally different targets. Finally, we show that the learning model can be reduced to a minimal model

where only the time constant of recovery from depression τ_{rec} needs to be learnt in order to obtain neurons firing at high or low frequency. Comparing this finding with the results from Carvalho and Buonomano [2011], we suggest that different parameters of the model describing STP might be related to different types of coding.

5.1 Methods

5.1.1 Single neuron model

Each neuron is modelled as in Carvalho and Buonomano [2011]: The sub-threshold dynamics of the electrical potential V_i of the generic neuron i are described by the equation:

$$\frac{dV_i}{dt} = -g_L V_i + \sum_{j=1,\, j\neq i}^{N} g_{ij}\,(E_{rev} - V_i)\,, \tag{5.1}$$

where $E_{rev} = 30\,mV$ is the reversal potential and $g_L = 0.1\,\mu S$ is the leak conductance - both quantities are equal and fixed for all neurons. $\{g_{ij}\}_{i,j=1,\ldots N}$ is the matrix of conductances and the generic element g_{ij} represents the conductance of the synapse going from neuron j to neuron i. Upon arrival of a presynaptic action potential elicited by neuron j, each of the conductances g_{ij} with $i = 1,\ldots N$, $i \neq j$ increases by a quantity w_{ij}, called effective synaptic efficacy, and decays exponentially back to zero with a fixed time constant $\tau_g = 10\,ms$, equal for all synapses:

$$\frac{dg_{ij}}{dt} = -\frac{g_{ij}}{\tau_g} + \sum_f w_{ij}\,\delta\left(t - t_j^f\right)\,, \tag{5.2}$$

where t_j^f is the f-th spike emitted by neuron j. The effective synaptic efficacy depends on both presynaptic and postsynaptic factors:

$$w_{ij} = r_{ij} u_{ij} A_{ij}\,, \tag{5.3}$$

where r_{ij} and u_{ij} are the presynaptic variables representing depression and facilitation in the STP model (see Subsection 5.1.2 below) and A_{ij} is the postsynaptic variable for the maximum synaptic strength (or absolute efficacy), which represents the maximum synaptic response (see Subsection 5.1.7). If $V_i(t) \geq 1\,mV$ a spike is elicited by neuron i and $V_i(t+dt)$ is set to 0 for the next $t_{ref} = 10\,ms$ (refractory period).

5.1.2 STP model

Short-term synaptic plasticity is described at each synapse through the evolution of two variables, r_{ij} and u_{ij}, representing the degree of depression and facilitation of the synapse connecting neuron j to neuron i. The time course of r_{ij} and u_{ij} is given by the following kinetic equations (Maass and Markram [2002]; Markram et al. [1998b]):

$$\frac{dr_{ij}}{dt} = \frac{1 - r_{ij}}{\tau_{rec_{ij}}} - \sum_f r_{ij} u_{ij} \delta\left(t - t_j^f\right) \tag{5.4}$$

$$\frac{du_{ij}}{dt} = \frac{U_{ij} - u_{ij}}{\tau_{facil_{ij}}} + \sum_f U_{ij} \left(1 - u_{ij}\right) \delta\left(t - t_j^f\right). \tag{5.5}$$

U_{ij}, $\tau_{rec_{ij}}$ and $\tau_{facil_{ij}}$ are the parameters of the model and they represent, respectively: Fraction of resources used by the first action potential, time constant of recovery from depression and time constant of synaptic facilitation. According with the derivation given in Subsection 2.3.2, the variables r_{ij} and u_{ij} correspond to precise biological quantities, respectively: occupancy of the release pool and probability of neurotransmitter release, both related to the presynaptic cell. Indeed, the presynaptic terminal contains a specilised area where vescicles filled with the neurotransmitters required for the signnal transmission are clustered. Calcium ions are the agents causing vescicles to migrate towards the presynaptic membrane, where they merge with the cell membrane releasing the neurotransmitters into the synaptic cleft. The arrival of a presynaptic action potential at the axon terminal triggers the opening of voltage gated calcium channels with subsequent influx of extracellular Ca^{++} by diffusion. This leads to an increase in the number of vescicles driven towards the cell membrane, and therefore in the number of released neurotransmitters.

A learning rule for STP has to allow changes to (at least one of) these parameters. At each synapse, the product of r_{ij} and u_{ij} determines the presynaptic efficacy.

5.1.3 STDP model

We use the triplet learning rule defined by Pfister and Gerstner [2006] with hard bounds: Maximum weights can only vary in the interval $[A_{min}, A_{max}]$. In this model, each neuron has two presynaptic variables m^1, m^2 and two postsynaptic variables o^1, o^2. In the absence of any activity, these variables exponentially decay towards zero with different time constants:

$$\tau_{m_1} \frac{dm_i^1}{dt} = -m_i^1 \quad \tau_{m_2} \frac{dm_i^2}{dt} = -m_i^2 \quad \tau_{o_1} \frac{do_i^1}{dt} = -o_i^1 \quad \tau_{o_2} \frac{do_i^2}{dt} = -o_i^2 \tag{5.6}$$

whereas when the neuron elicits a spike they increase by 1:

$$m_i^1 \to m_i^1 + 1 \quad m_i^2 \to m_i^2 + 1 \quad o_i^1 \to o_i^1 + 1 \quad o_i^2 \to o_i^2 + 1\,. \tag{5.7}$$

Then, assuming that neuron i fires a spike, the STDP implementation of the triplet rule can be written as follows:

$$\begin{cases} \Delta A_{ji}^{STDP} = -\gamma o_j^1(t)\left[A_2^- + A_3^- m_i^2(t-\varepsilon)\right] \\ \Delta A_{ij}^{STDP} = +\gamma m_j^1(t)\left[A_2^+ + A_3^+ o_i^2(t-\varepsilon)\right] \end{cases} \tag{5.8}$$

where γ is the learning rate, ε is an infinitesimal time constant to ensure that the values of m_i^2 and o_i^2 used are the ones right before the update due to the spike of neuron i, and A_{ij} is the maximum strength of the connection from j to i. Values of STDP amplitudes are taken from Pfister and Gerstner [2006] and are listed in Tab. 5.1.

In order to set A_{min} we note that if the maximum weights connecting the input neurons to a specific output neuron all collapse to zero in the low firing rate regime, then, in the subsequent high firing rate regime, inputs were not able to "wake up" this neuron: It remained almost silent all the time. To avoid this, we set $A_{min} = 10^{-3}$. With such a small value we can still apply the symmetry measure (Esposito et al. [2014]), which assumes $A_{min} = 0$, see Subsection 5.1.9, to evaluate the symmetry of the network.

5.1.4 Learning task

Neurons are divided into different populations, each of them is required to fire at one of the two target firing rates: $30\,Hz$ (*high*) or $5\,Hz$ (*low*). To allow the populations to reach their target rate, both short- and long-term plasticity parameters are adapted via error-driven learning (see Subsection 5.1.6) and, in addition, the maximum synaptic strength is shaped by the STDP triplet rule (see Subsection 5.1.3).

5.1.5 Input signal and input neurons

In all simulations, the input signal is delivered only to a subset of neurons in the network, which we call input neurons N_{in}. Each of these neurons receives a pulse-like stimulus with a fixed frequency $\nu_{in} = 10\,Hz$, whose amplitude ($2\,mV$) is chosen to always elicit an action potential in the corresponding input neuron. The stimulus delivery, however, is not synchronous across the input neurons, but it follows a *sequential protocol*: Neurons are stimulated one after another with a fixed time delay t_{delay} and in a fixed order. We

choose $t_{delay} = (\nu_{in} N_{in})^{-1}$ so that neurons that belong to input cyclically receive a stimulus. To further explain this, one may imagine labelling the neurons depending on the order they receive the stimulus, and therefore on the firing order, then have the firing pattern $N_1, N_2, N_3, \ldots, N_{N_{in}}, N_1, N_2, N_3, \ldots, N_{N_{in}}, N_1, \ldots$, with each pair of adjacent spikes being separated by a time interval of t_{delay}. We can think of the N_{in} neurons as if they are organised in a ring and the stimulus as a cyclically travelling wave across this ring. To include the effect of noise, a random Gaussian variable with zero mean and standard deviation equal to $0.1\, t_{delay}$ is added to the firing times. The magnitude of the standard deviation is such that there is no inversion in the firing order. With this construction, the stimulus delivered to input neurons can be thought as generated by an external (not explicitly simulated) population of neurons where each external neuron projects only onto one corresponding input neuron.

Note that, by construction, in the absence of any other signal, the firing pattern of the input neurons reflects that of the stimulus. This means that the external signal implicitly fixes a level of minimum activation for the N_{in} neurons: Their firing rate cannot be smaller than ν_{in}. Due to this constraint, the input neurons, despite being free to change their parameters according to STP learning rules (see Subsection 5.1.6), are not totally free to regulate their firing activity, which may prevent them from effectively fulfilling the task. The rest of the neurons, instead, are totally free to adapt their activity and are called output neurons. For these reasons, we read out the interesting quantities only from output neurons (we refer to Results and to Fig. 5.1A, 5.4A for more details on the architecture).

5.1.6 Error-driven learning rule for STP

The task can be formulated as an optimisation problem where neurons regulate their own activity in order to minimise the objective function defined as:

$$E = \left(\frac{\nu_{targ} - \langle \nu \rangle}{\nu_{lim}} \right)^2 , \tag{5.9}$$

where ν_{lim} is the maximum allowed frequency due to the refractory period ($\nu_{lim} = 1/t_{ref}$), ν_{targ} is the target firing rate and $\langle \nu \rangle$ is the mean firing rate across a single population. To calculate firing rates of single neurons ν_i we use an exponential moving average with time constant $\tau_\nu = 1\, s$:

$$\tau_\nu \frac{d\nu_i}{dt} = -\nu_i + \hat{\nu}_i \tag{5.10}$$

where $\hat{\nu}_i$ is the current firing rate, which basically reflects if neuron i has fired ($\hat{\nu}_i = 1\,Hz$) or not ($\hat{\nu}_i = 0\,Hz$). The population mean firing rate is therefore:

$$\langle \nu \rangle = \frac{1}{N_{pop}} \sum_{i=1}^{N_{pop}} \nu_i \tag{5.11}$$

with N_{pop} being the size of the population.

By following a standard procedure, learning rules can be derived from Eq. (5.9) by applying the gradient descent method (Hertz et al. [1991]). Since the task is not based on single neurons but it involves an entire population, we use a mean-field approach for the derivation of the learning rules. Therefore, from now on in this section, we switch from the above single neuron notation to mean-field variables, by dropping the ij indices. It is worth noting that in our formulation the target is achieved not by directly acting on the firing rates, but by tuning the STP parameters, which in turn affects the firing itself. Therefore, $\langle \nu \rangle = \langle \nu \rangle \left(U, \tau_{rec}, \tau_{facil}\right)$ and by using the chain rule we can formally write the following update rule for each parameter p:

$$\Delta p = -\eta_p \frac{\partial E}{\partial p} = -\eta_p \frac{\partial E}{\partial \langle \nu \rangle} \frac{\partial \langle \nu \rangle}{\partial p} = 2\eta_p \frac{\nu_{targ} - \langle \nu \rangle}{\nu_{lim}^2} \frac{\partial \langle \nu \rangle}{\partial p}, \qquad p = U, \tau_{rec}, \tau_{facil} \tag{5.12}$$

where η_p is the learning rate, which in principle could be different for each parameter. The form of the function $\langle \nu \rangle \left(U, \tau_{rec}, \tau_{facil}\right)$ can be derived with a semi-heuristic procedure based on mean field equations for a population with short-term plastic synapses, which in turn can be obtained by averaging the STP single neuron equations, Eq. (5.4),(5.5). In doing so, to resolve terms like $\left\langle r_{ij} u_{ij} \delta\left(t - t_j^f\right)\right\rangle$, we make the following two assumptions (Barak and Tsodyks [2007]; Tsodyks et al. [1998]): *i)* each neuron fires a Poisson train; *ii)* there is no statistical dependence between the variables $r(t)$ and $u(t)$ and the probability of spike emission at time t. The first assumption guarantees that the probability that a neuron fires at time t depends only on the instantaneous firing rate and not on the timing of previous spikes, and that $r(t)$ and $u(t)$, conversely, are a function only of the spike arrival times prior to the current time and do not dependent on the probability of a spike at time t. Strictly speaking, having refractoriness breaks down the Poisson assumption; however, since the maximum firing rate is $30\,Hz$, the refractory period ($10\,ms$) is always much shorter than the minimum inter-spike interval ($\sim 33\,ms$). Hence, we can consider the Poisson assumption as a good approximation, and therefore write: $\left\langle r_{ij} u_{ij} \delta\left(t - t_j^f\right)\right\rangle = \langle r_{ij} u_{ij} \rangle \left\langle \delta\left(t - t_j^f\right)\right\rangle$. Concerning the second assumption, it is always valid if there is no facilitaiton, as $u(t) = U$ is a constant. If facilitation is included, instead, $r(t)$ and $u(t)$ are statistically dependent. However, it can be shown that the relative error $|\langle r_{ij} u_{ij} \rangle - \langle r_{ij} \rangle \langle u_{ij} \rangle| / \langle r_{ij} \rangle \langle u_{ij} \rangle$ can exceed 10%

only at very short values of τ_{facil} and high values of U at which the model does not exhibit facilitating behavior anymore (Tsodyks et al. [1998]). Hence, our second assumption can be considered a valid approximation.

Following Vasilaki and Giugliano [2014] (and using the same notation whenever possible), we thus introduce the mean-field variables u, x, U and A, respectively describing facilitation, depression, synaptic utilisation and maximum strength, and we assume a threshold-linear gain function between input mean current h and output mean firing rate $\langle \nu \rangle = a\left[(h-\vartheta)\right]_+$, for some constants a, ϑ. We can then finally write the dynamic mean-field equations for a population of neurons recurrently connected by short-term synapses as follows (Chow et al. [2005]):

$$\begin{cases} \tau \dot{h} = -h + Aux \langle \nu \rangle + I_{ext} \\ \dot{x} = \frac{1-x}{\tau_{rec}} - ux \langle \nu \rangle \\ \dot{u} = \frac{U-u}{\tau_{facil}} + U(1-u) \langle \nu \rangle \end{cases} \tag{5.13}$$

where I_{ext} is the mean external current and τ is a decaying constant. By imposing equilibrium conditions, $\dot{h} = \dot{x} = \dot{u} = 0$, and combining the resulting equations, we can finally write:

$$h = F\left(\langle \nu \rangle_h\right) = \frac{AU\left(\langle \nu \rangle^{-1} + \tau_{facil}\right)}{\langle \nu \rangle^{-2} + \langle \nu \rangle^{-1} U \tau_{facil} + \langle \nu \rangle^{-1} U \tau_{rec} + U \tau_{facil} \tau_{rec}} + I_{ext} \tag{5.14}$$

Now we observe that by taking the limit $h \to \infty$ in $F\left(\langle \nu \rangle_h\right)$ we obtain an upper bound for the maximum allowed firing rate $\langle \nu \rangle \leq \frac{A}{\tau_{rec}} + I_{ext}$ (for more details see Vasilaki and Giugliano [2014]). We can heuristically turn the above inequality into an equality:

$$\langle \nu \rangle = \frac{A}{\tau_{rec}} + I_{ext} \tag{5.15}$$

so as by plugging Eq. (5.15) into Eq. (5.12) we can finally obtain an explicit form for the learning rule. In particular, since only one of the three parameters appears in Eq. (5.15), we have a single rule for τ_{rec} only:

$$\Delta \tau_{rec} = -2\eta_{\tau_{rec}} \left(\nu_{targ} - \langle \nu \rangle\right) \frac{A}{\nu_{lim}^2 \tau_{rec}^2} \tag{5.16}$$

Then, according to the above derivation, the only parameter that needs to be learnt is τ_{rec}. Here we adopt the view (Chow et al. [2005]; Markram et al. [1998b]; Thomson [2000]; Tsodyks and Markram [1997]) that facilitation/depression corresponds to small/large values

of τ_{rec} and U as well. Therefore, assuming that they apparently play a similar role, we can heuristically take a similar dependence of $\langle \nu \rangle$ upon U: $\langle \nu \rangle = \frac{A}{U} + I_{ext}$, which leads us to a similar learning rule:

$$\Delta U = -2\eta_U \left(\nu_{targ} - \langle \nu \rangle \right) \frac{A}{\nu_{lim}^2 U^2} \tag{5.17}$$

With the same heuristic argument we can also write down a relation involving τ_{facil}. Indeed, it is well know that facilitation/depression corresponds to large/small values of τ_{facil}, so we can hypothesise a linear relation, also including the dependence on the maximum strength for similarity with the other parameters. Thus, $\langle \nu \rangle \propto A\tau_{facil} + I_{ext}$, which gives the following learning rule:

$$\Delta \tau_{facil} = 2\eta_{\tau_{facil}} \left(\nu_{targ} - \langle \nu \rangle \right) \frac{A}{\nu_{lim}^2} \tag{5.18}$$

Finally, based on the fact that A turns out to appear in Eq. (5.15), and supported by experimental results showing an interaction between STP and STDP (Markram et al. [1997]; Sjöström et al. [2003]), we can also introduce a STP-dependent learning rule for the maximum synaptic strength:

$$\Delta A^{STP} = -\eta_A \frac{\partial E}{\partial A} = -\eta_A \frac{\partial E}{\partial \langle \nu \rangle} \frac{\partial \langle \nu \rangle}{\partial A} = 2\eta_A \left(\nu_{targ} - \langle \nu \rangle \right) \frac{1}{\nu_{lim}^2 \tau_{rec}} . \tag{5.19}$$

This synaptic modification clearly does not substitute the traditional STDP, since the two rules come from different mechanisms. Rather, we assume they both contribute to maximum weights changes, as we show below.

5.1.7 Single neuron learning framework: Combining STDP and STP learning models

Eq. (5.16)-(5.19) are mean field learning rules for the four parameters $\tau_{rec}, U, \tau_{facil}, A$. It is straightforward to turn them into single neuron online learning rules. From now on, we return to a single neuron notation. Similarly to STDP, we hypothesise that modifications of STP are triggered by postsynaptic events: Every time neuron i elicits a spike, its current firing rate is updated as well as the mean population firing rate. Neuron i can therefore backwards

regulate its incoming synapses, through the following set of equations:

$$\Delta\tau_{rec_{ij}} = -2\eta_{\tau_{rec_{ij}}}\left(\nu_{targ} - \langle\nu\rangle\right)\frac{A_{ij}}{\nu_{lim}^2\tau_{rec_{ij}}^2} \tag{5.20}$$

$$\Delta U_{ij} = -2\eta_{U_{ij}}\left(\nu_{targ} - \langle\nu\rangle\right)\frac{A_{ij}}{\nu_{lim}^2 U_{ij}^2} \tag{5.21}$$

$$\Delta\tau_{facil_{ij}} = 2\eta_{\tau_{facil_{ij}}}\left(\nu_{targ} - \langle\nu\rangle\right)\frac{A_{ij}}{\nu_{lim}^2} \tag{5.22}$$

$$\Delta A_{ij}^{STP} = 2\eta_{A_{ij}}\left(\nu_{targ} - \langle\nu\rangle\right)\frac{1}{\nu_{lim}^2\tau_{rec_{ij}}}\,. \tag{5.23}$$

The firing event of the neuron i also triggers STDP, according with Eq. (5.8). This contribution sums up with the above STP-dependent change, so as the total modification of the maximum synaptic strength is:

$$A_{ij} \longrightarrow A_{ij} + \Delta A_{ij}^{tot}, \qquad \Delta A_{ij}^{tot} = \Delta A_{ij}^{STDP} + \Delta A_{ij}^{STP}\,. \tag{5.24}$$

Note that when we converted mean field population equations into single neuron equations we kept the population mean firing rate $\langle\nu\rangle$, instead of turning it into the single rate ν_i. This is because the task is defined at a population level. Learning rates of the three STP parameters are chosen to be equal and error-dependent:

$$\eta_{p_{ij}} = \bar{\eta}\left(1 + \frac{\nu_{targ} - \langle\nu\rangle}{\nu_{lim}}\right)^2, \qquad p = U, \tau_{rec}, \tau_{facil}, \tag{5.25}$$

with $\bar{\eta} = 0.1$. The learning rate for maximum synaptic strength, instead, is fixed in time and it is the same as the one used for STDP, $\eta_{A_{ij}} \equiv \gamma$.

Now we have four single neuron rules for the STP learning model, plus an equation for STDP and an equation for combining the different rules for the maximum synaptic strength. All these six rules together, Eq. (5.8),(5.20),(5.21),(5.22),(5.23),(5.24) form a complete learning scheme for each neuron, which is implemented in our simulations. These rules are now local, since their computation takes place separately in each neuron, but receive a global signal encoding for the task performance error.

5.1.8 Investigation of different rule combinations

In the Results section we consider different learning mechanisms: In addition to STDP, that is crucial for the formation of motifs (Vasilaki and Giugliano [2012, 2014]), different combinations of the four STP rules are taken into account while the remaining parameters

are kept fixed. At first we allow only two parameters to change: *i)* τ_{rec}, because Eq. (5.15) implies that for high frequencies this is the only critical parameter for adapting the firing rate of the population, and *ii)* U, since it was a key parameter adopted in the work in Carvalho and Buonomano [2011]. Then, we introduce the STP-dependent rule on the maximum synaptic strength, Eq. (5.23), with the view to observe a more stable learning process. Following this, we also include τ_{facil} in the learning scheme for a full parameter adaptation (*full model*) and finally we investigate the minimal number of parameters that needs to be adapted (*minimal model*), based on Eq. (5.15). Looking for other parameter combinations might not be meaningful, as Eq. (5.15) indicates the key parameters that are involved in changing the mean firing of the population.

5.1.9 Connectivity analysis

To reveal the type of connectivity in the output population, we use a symmetry index defined as a measure of the symmetry of the connectivity matrix W (Esposito et al. [2014]):

$$s = 1 - \frac{2}{N(N-1) - 2M} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{\left|A_{ij} - A_{ji}\right|}{A_{ij} + A_{ji}}. \tag{5.26}$$

Here M is the number of instances where both A_{ij} and A_{ji} are zero, i.e. there is no connection between two neurons. Since in our case connections are bounded in the interval $\left[10^{-3}, 1\right]$, $M = 0$ all the time. Eq. (5.26) is able to capture the presence of global non-random structures in a network, returning a value included in $[0, 1]$. Values of s close to 1 reflect the presence of a global bidirectional motif, whereas when s approaches 0, a unidirectional motif is emerging. Note that, in order to apply the measure Eq. (5.26), we assume that the lower bound for connections is 0. However, the choice of a small value such as 10^{-3} does not affect the measure.

Data Sharing We provide the scripts that were used to construct the main figures of the paper in the ModelDB database, accession number: 169242.

Symbol	Description	Value
N	Number of total neurons	$\{40, 80\}$
N_{in}	Number of input neurons	$\{30, 60\}$
N_{out}	Number of output neurons	10
E_{rev}	Reversal potential	30 mV
g_L	Decay constant of neuron potential	0.1 μ S
τ_g	Decay constant of synaptic conductances	10 ms
V_{thr}	Threshold potential for spike emission	1 mV
t_{ref}	Refractory period	10 ms
ν_{in}	Input firing rate	10 Hz
ν_{targ}	Output target firing rate	$\{2, 20\}$ Hz
τ_{facil}^{min}	Facilitation time constant - minimum value	1 ms
τ_{facil}^{max}	Facilitation time constant - maximum value	900 ms
τ_{rec}^{min}	Depression time constant - minimum value	100 ms
τ_{rec}^{max}	Depression time constant - maximum value	900 ms
U^{min}	Synaptic utilisation - minimum value	0.05
U^{max}	Synaptic utilisation - maximum value	0.95
$\bar{\eta}$	Fixed learning rate for U, τ_{rec}, τ_{facil}	0.1
A_2^+	Amplitude of maximum weights change - pair term in LTP	4.6×10^{-3}
A_3^+	Amplitude of maximum weights change - triplet term in LTP	9.1×10^{-3}
A_2^-	Amplitude of maximum weights change - pair term in LTD	3.0×10^{-3}
A_3^-	Amplitude of maximum weights change - triplet term in LTD	7.5×10^{-9}
τ_{m_1}	Decay constant of presynaptic indicator m_1	16.8 ms
τ_{m_2}	Decay constant of presynaptic indicator m_2	575 ms
τ_{o_1}	Decay constant of postsynaptic indicator o_1	33.7 ms
τ_{o_2}	Decay constant of postsynaptic indicator o_2	47 ms
A_{min}	Lower bound for maximum synaptic weights	0.001
A_{max}	Higher bound for maximum synaptic weights	1
γ	Learning rate for STDP and for STP-dependent Δw	$\{1, 2\}$
dt	Discretisation time step	1 ms

Table 5.1 Parameters used in simulations. STDP parameters are as in the nearest-spike triplet-model, described in Pfister and Gerstner [2006].

5.2 Results

5.2.1 Single population with a time-varying task: A continuum between facilitation and depression

First, we apply our learning model to a specific task demonstrating how synapses can change their behaviour driven by an external feedback signal. The problem we study is

simple: A population of neurons is presented with a stimulus and is required to produce a certain output as a response to that stimulus. Once the learning has been successful, for the same input signal the desirable output changes. In other words, neurons are trained to respond to a change in the associative paradigm (*inverted associative learning problem*), that can be due to, for instance, changes in the environmental conditions.

Let us give a concrete example of an inverted associative learning problem, taken from Asaad et al. [1998]. In their work, the authors trained monkeys to associate visual stimuli (pictures) with delayed saccadic movements, left or right, with associations being reversed from time to time. Monkeys had to go beyond learning a single cue-response association: They are required to learn to associate, on alternate blocks, two cue objects with two different saccades. In other words, after having learned the relation $\{$*object A, go right*$\}$, and $\{$*object B, go left*$\}$, the associations were reversed such that now they needed to learn $\{$*object A, go left*$\}$ and $\{$*object B, go right*$\}$.

Similar to the Asaad et al. [1998] experiment, we assume a binary problem, i.e. environmental conditions can change only between two states, and we measure the neurons' activity in terms of firing rate. This means that neurons are initially asked to fire at some rate and, after learning this task, they are asked to fire at a different rate, while keeping the same input signal all the time. Thus, the problem we defined is a simpler version of the monkey experiment, with only a single input. In order to train the neurons on the current associative paradigm, an external global signal is required, that can be considered as an error signal (see Methods 2.6 and 2.7).

Problem description and network architecture

We created a learning network of $N = 40$ conductance-based integrate-and-fire neurons (see Methods 2.1) all to all connected. Synaptic connections are modified by the STDP triplet rule (Pfister and Gerstner [2006]) and short-term plasticity is implemented by using the Tsodyks and Markram model (TM model) described in Maass and Markram [2002]; Markram et al. [1998b].

Fig. 5.1A shows the network architecture, composed by two non-overlapping regions: A *blue* one with $N_{in} = 30$ neurons receiving the input signal and a *red* one with $N_{out} = 10$ neurons from which we read out the quantities of interest. Note that for clarity, only a few neurons (*black circles*) and connections (*black arrows*) are drawn. The network is therefore formed by two functionally distinct populations, with the input population delivering the stimulus to the output one. Recursive connections are present within each population and across populations, and they are all plastic, in the sense of both long-term and short-term plasticity. We refer to this architecture as a first or single population scenario.

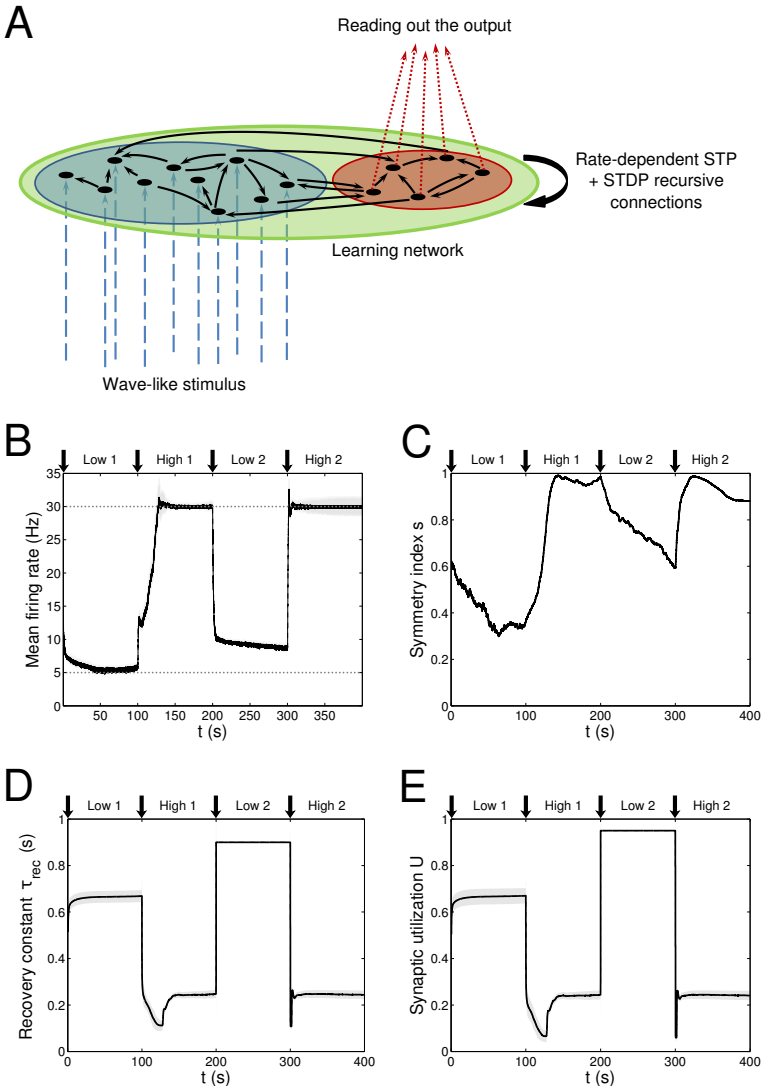

Figure 5.1 Single population scenario: Network architecture and activity, connectivity and STP parameters adaptation in the output population with (U, τ_{rec}) learning scheme (*Part 1*). A Architecture. The learning network (*green*) is divided into an input region (*blue*) and an output region (*red*). Connections (*black arrows*) are all-to-all and obey both Spike-Timing Dependent Plasticity and rate-dependent short-term plasticity. Input neurons receive an external wave-like stimulus (*blue dashed arrows*). **B** Mean firing rate of the output population. *Shaded area* represents standard deviation, *horizontal dotted grey lines* show the two target firing rates (high $= 30\,Hz$, low $= 5\,Hz$) and *vertical black arrows* mark the onset of the four dynamic phases alternating the target according to the sequence low-high-low-high. **C** Symmetry measure applied on the connectivity of the output population. In accordance to the target, connectivity switches between unidirectionality (low values) and bidirecitonality (high values). **D**, **E** Mean values of the recovery time constant τ_{rec} and synaptic utilisation U for the synapses projecting onto the output neurons. We observe depression (high values) at low firing rates and facilitation (low values) at high firing rates.

The input neurons are stimulated one after the other, following a *sequential protocol*, and approximately with the same rate, $\nu_{in} = 10\,Hz$. The amplitude of the stimulus is such that input neurons release a spike every time they receive an input (see Methods 2.5). This external source can be thought as an additional population of neurons, which we are not simulating here, where each "external" neuron is connected only with a corresponding neuron in the input population by means of a fixed feedforward connection (*blue dashed arrows*).

We hypothesise that the whole learning network (*green* region in Fig. 5.1A) is presented with a sequence of two tasks while the stimulus pattern is kept fixed. The tasks are firing low ($5\,Hz$) and firing high ($30\,Hz$) and the sequence is *low-high-low-high*. Therefore, neurons have to repeatedly learn a new association and forget the previous one in a dynamic context divided in four phases of $t_{ph} = 100\,s$. We refer to them as: Low 1, high 1, low 2, high 2. As discussed at the beginning of this section, this picture is inspired by a typical inverted associative learning problem: Considering the monkey experiment from Asaad et al. [1998] as a metaphor, our scenario provides a simplified version, where instead of having two different inputs, *object A* and *object B*, we have a single input. Indeed, we can think we are presenting the network with only *object A* and while doing this we switch the target association between the two states *go right* and *go left*, which correspond to our low and high firing rate targets. We call the desirable context-dependent target rate, ν_{targ}. As described in Methods, the difference between ν_{targ} and the current firing rate of each population $\langle \nu \rangle$ is the error signal that, according which our rate-dependent STP causes synapses to adapt their activity.

In all simulations, single neuron parameters $\left\{U, \tau_{rec}, \tau_{facil}, w\right\}_{ij}$ are initially drawn from uniform distributions (for $i \neq j$), respectively in $[0.05, 0.95]$, $[100, 900]\,ms$, $[1, 900]\,ms$, $\left[10^{-3}, 1\right]$. Synaptic variables are initialised at their equilibrium values, i.e. $r_{ij} = 1$ and $u_{ij} = U_{ij}$. All the simulations in this subsection use $\gamma = 1$ for the high rate regime and $\gamma = 2$ for the low rate regime. Values of the parameters are listed in Tab. 5.1.

Learning U and τ_{rec}

We initially studied the problem with a learning scheme involving U and τ_{rec} only, Eq. (5.20), (5.21), so there is no additional change in maximum synaptic strengths due to STP. Indeed, due to Eq. (5.15) and Carvalho and Buonomano [2011], we wanted to test the hypothesis that U and τ_{rec} are the only crucial parameters that need to be learnt for adapting the firing rate of a population. The results are displayed in Fig. 5.1B-E, with *vertical black arrows* marking the beginning of each of the four phases, and in Fig. 5.2.

Fig. 5.1B shows the average firing rate of the N_{out} neurons, with *shaded area* being the standard deviation. Target firing rates are show with *grey dotted lines*. The adaptation to

the new target is fast, except during the *low 2* phase, when we switch from high to low rate, where an initial fast decrease of the firing rate is followed by a much slower adaptation. Despite the fact that neurons do not reach the target rate during this phase, we observe a monotonically decreasing activity which would eventually stabilise at $5Hz$ if we were allowing the simulation to run for longer. The reason for this double slope adaptation will be further discussed now.

Fig. 5.1C shows the evolution of the symmetry index (see Methods 2.9). At the beginning, the value reflects the randomness in the connections (the mean value of s for a network with uniform random connections is indeed $\simeq 0.614$, see Esposito et al. [2014]), whereas, as learning takes place, we observe the development of unidirectional (low values of s) and bidirectional (high values of s) motifs, depending on the set target. This can also be formalised by applying the p-value hypothesis test obtained by using mean and variance of s on a completely random network with uniform distribution of connections (Esposito et al. [2014]). P-values are shown in Tab. 5.2. We, again, observe rather slow dynamics during the *low 2* phase that, within the fixed simulation time, prevent the system from reaching a clear connectivity configuration. However, the trend of s clearly shows that the connectivity within the output population is approaching unidirectionality.

Fig. 5.1D,E depicts the time course of the recovery time constant τ_{rec} and synaptic utilisation U averaged across the output neurons, with *shaded area* representing standard deviation. Both parameters oscillate between high values, which correspond to depressing behaviour, and low values, that indicate facilitation. Note that the dynamics of τ_{rec} and U is fast in all phases, the third included. This is not surprising since STP is a fast process and leads to fast adaptation of its parameters. As a result, neurons' response to a change in the target rate takes place in a short time. However, during the *low 2* phase, synaptic parameters saturate before the neurons could fulfil the task, with STDP being the only remaining mechanism through which the output population can regulate its own activity. This results in a much slower decrease towards the target rate for two reasons: *i)* STDP by itself acts on much longer time scales, *ii)* switching from high to low rate is the most challenging part of the entire sequence of tasks due to the saturation of the maximum weights in the previous *high 1* phase, which slows down the process even further.

Fig. 5.2 provides additional evidence of the alternation between the two different synaptic behaviours. Plots are organised in five rows, with each row displaying information from a different phase of the simulation. Panel A shows the initial uniform condition, panel B the end of *low 1* phase, etc. For each stage, we draw the histograms of recovery time constant (*Column 1*) and synaptic utilisation (*Column 2*). According to the narrow standard deviation observed in Fig. 5.1D,E, distributions peak around extreme values, reflecting two different,

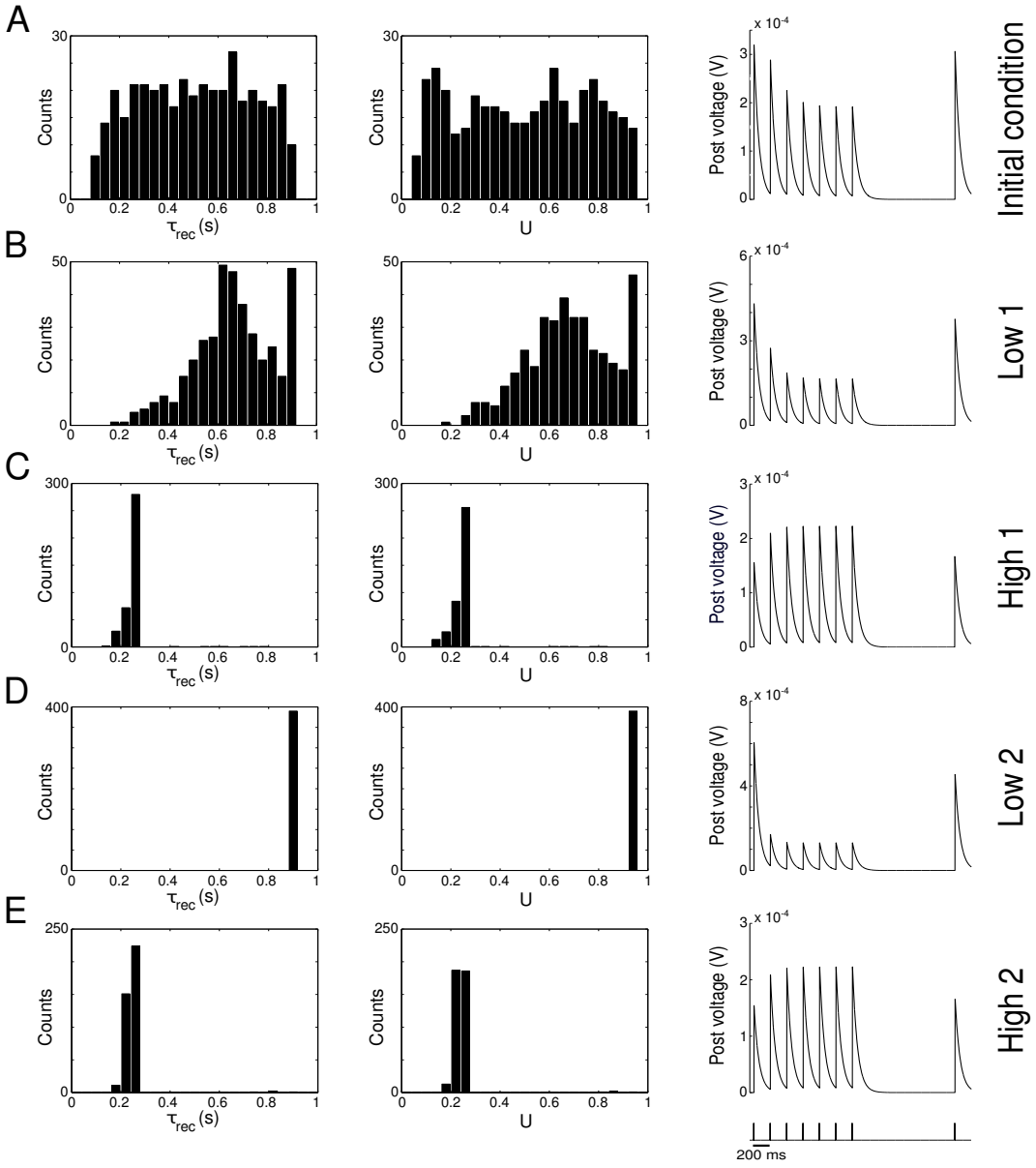

Figure 5.2 Single population scenario: STP parameters distribution during adaptation with (U, τ_{rec}) learning scheme (*Part 2*). Different phases of the dynamics are represented. **A** Initial (uniformly random) condition. **B** End of *low 1* phase (target rate is $5\,Hz$). **C** End of *high 1* phase ($30\,Hz$). **D** End of *low 2* phase ($5\,Hz$). **E** End of *high 2* phase ($30\,Hz$). *Columns 1, 2* Histograms of recovery time constant and synaptic utilisation of the synapses projecting onto the output neurons. Low values indicates facilitation whereas high values suggest depression. *Column 3* Single synapse traces obtained with the TM model by applying a $5\,Hz$ stimulus. Synaptic parameters used are mean values obtained from the distributions drawn in panels A and B. Synapses display a clear alternation between depressing and facilitating behaviour.

synaptic behaviours. *Column 3* in Fig. 5.2, displaying the single synapse traces obtained with a TM model, demonstrates the corresponding behaviours: At the end of the phases where neurons are required to fire low we observe a typical depressing response, whereas at the end of the high rate regimes synapses show a typical facilitating trace. To generate the traces, we used a $5\,Hz$ signal to stimulate a synapse with a parameters given by the mean values obtained from the corresponding histograms. Note that the synaptic trace for the initial condition, i.e. before learning shapes the parameters, already shows depression, which explains why the distributions of τ_{rec} and U at the end of the *low 1* phase are much broader than in the following phases.

Altogether, the four panels B-E in Fig. 5.1 and the five panels A-E in Fig. 5.2 show that the properties and activity of the output population oscillate between two states and that the desirable structure is formed depending on the target rate. In particular, we observe that neurons that fire at low frequency turn their synaptic properties into depressing and the connections formed are mostly unidirectional. Conversely, when the target rate is set at a high frequency, neurons develop facilitating synapses and bidirectional connections.

STP-dependent modification of A enhances performance

Given the speed convergence issue in the *low 1* phase, we introduced an additional learning mechanism, i.e. the STP-dependent rule for A, Eq. (5.23), (5.24). Indeed, this mechanism provides an additional way, besides the STDP, for regulating the long-term plastic synapses. In all the other aspects, the model remains as above.

Fig. 5.3 shows simulation results, with panels A-D depicting the same quantities as panels B-E in Fig. 5.1 (symbols as before). A direct panel-by-panel comparison shows that the results are very similar, meaning that with this new learning configuration the output population also learns to adapt its synaptic properties in order to fulfil the current task, with subsequent motifs formation. As expected, due to the additional learning rule for A, the dynamics are faster: In particular, during the *low 1* phase, neurons reach the target rate within the simulation time, and the value of the symmetry measure is much lower than before, confirming the formation of a unidirectional motif; compare with Fig. 5.1C, 5.3B and see Tab. 5.2. Note that the adaptation of the STP parameters is also faster, as they depend on the current value of the maximum synaptic strength. Thus, the STP-dependent modification of A improves the overall performance and introduces an interesting link between STP and STDP.

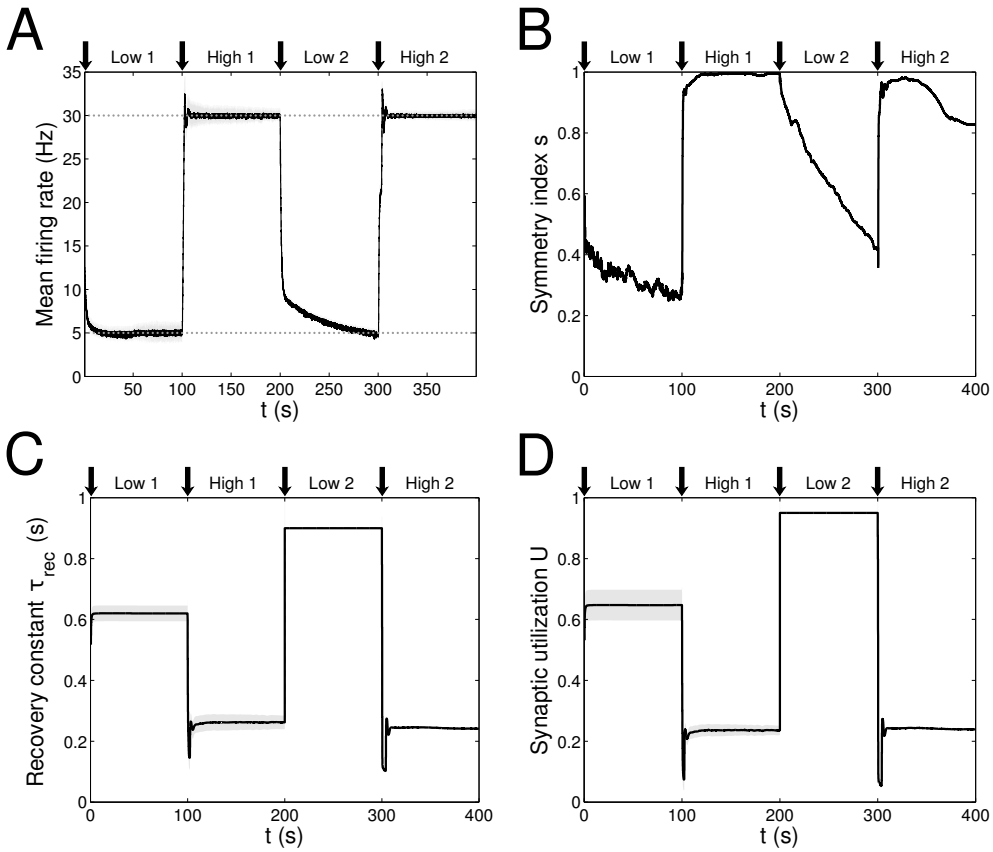

Figure 5.3 Single population scenario: Activity, connectivity and STP parameter adaptation in the output population with (U, τ_{rec}, A) learning scheme. **A** Mean firing rate of the output population. *Shaded area* represents standard deviation, *horizontal dotted grey lines* show the two target firing rates (high $= 30Hz$, low $= 5Hz$) and *vertical black arrows* mark the onset of the four dynamic phases alternating the targets in the sequence low-high-low-high. **B** Symmetry measure applied on the connectivity of the output population. In accordance with the target rate, connectivity switches between unidirectionality (low values) and bidirecitonality (high values). **C**, **D** Mean values of recovery time constant τ_{rec} and synaptic utilisation U for the synapses projecting onto the output neurons. We observe depression (high values) at low firing rates and facilitation (low values) at high firing rates. Compared to Fig. 5.1, we observe an improvement in the overall performance due to the inclusion of the STP-dependent modification of A.

5.2.2 Two populations with a different task: Synaptic differentiation

Now we consider a different scenario, which we refer to as the second or double population scenario. The two tasks associated with low and high targets are now simultaneously active and must be learnt by different populations, interacting via lateral connections and receiving the same stimulus source. Reasons are multiple: We want to investigate if our model allows to contemporary encode both associative paradigms, without the need of forgetting one of the two. In addition, we want to study the possibility that target-specific STP

Phase	(τ_{rec}, U) scheme		(τ_{rec}, U, A) scheme	
	s	p-value	s	p-value
Low 1 (0-500 s)	0.36	1.82×10^{-9}	0.28	1.13×10^{-15}
High 1 (500-1000 s)	0.98	5.47×10^{-19}	0.99	8.73×10^{-20}
Low 2 (1000-1500 s)	0.59	6.47×10^{-1}	0.41	2.19×10^{-6}
High 2 (1500-2000 s)	0.88	1.50×10^{-10}	0.82	3.63×10^{-7}

Table 5.2 Symmetry measure and p-value for the single population scenario with $\{\tau_{rec}, U\}$ and $\{\tau_{rec}, U, A\}$ learning schemes. *Column 1:* The four phases of dynamics with the corresponding simulation time. *Columns 2,3:* Symmetry measure and p-value for the adaptation of $\{\tau_{rec}, U\}$. Values are computed at end of each phase and by considering output neurons only. *Columns 4,5:* Same as columns 2 and 3 except for the adaptation of $\{\tau_{rec}, U, A\}$.

emerges as a consequence of the target-dependent learning rules we chose for our model. In particular, we want to test whether our model is able to reproduce existing experimental data, specifically that appearing in Table 1, from the paper by Wang and his collaborators (Wang et al. [2006]).

Synaptic groups	$\tau_{rec}(ms)$	$\tau_{facil}(ms)$	U	τ_{rec}/τ_{facil}	τ_{rec}/τ_{facil} as in Wang	Wang's subtypes
$\wp_1^{out} \bigcup \wp_2^{out} \rightarrow \wp_1^{out}$	310 ± 11	733 ± 17	0.27 ± 0.01	0.42	0.38	E1
$\wp_1^{out} \rightarrow \wp_1^{out}$	260 ± 5	833 ± 13	0.25 ± 0.01	0.31	0.34	E1a
$\wp_2^{out} \rightarrow \wp_1^{out}$	356 ± 19	643 ± 27	0.29 ± 0.01	0.55	0.43	E1b
$\wp_2^{out} \bigcup \wp_1^{out} \rightarrow \wp_2^{out}$	550 ± 14	440 ± 19	0.55 ± 0.02	1.25	39.47	E2
$\wp_2^{out} \rightarrow \wp_2^{out}$	595 ± 16	436 ± 26	0.61 ± 0.02	1.36	76.88	E2a
$\wp_1^{out} \rightarrow \wp_2^{out}$	510 ± 23	443 ± 28	0.50 ± 0.03	1.15	25.55	E2b

Table 5.3 Types and subtypes of excitatory synapses between the two output populations in the full model $\{\tau_{rec}, U, \tau_{facil}, A\}$. *Column 1:* Synaptic groups. For instance $\wp_1^{out} \bigcup \wp_2^{out} \rightarrow \wp_1^{out}$ includes all synapses from both output populations, $\wp_1^{out}$ and $\wp_2^{out}$, to the output population firing high, $\wp_1^{out}$. *Columns 2,3,4:* Mean values of STP parameters τ_{rec}, τ_{facil}, U. As in Wang et al. [2006], we provide the results in the form $mean \pm s.m.e.$. *Column 5:* Ratio between the two time constants, τ_{rec}/τ_{facil}, in our simulation. *Column 6:* For a direct comparison, we provide the values of τ_{rec}/τ_{facil} as in Wang et al. [2006]. *Column 7:* Mapping of our subtypes onto Wang's subtypes.

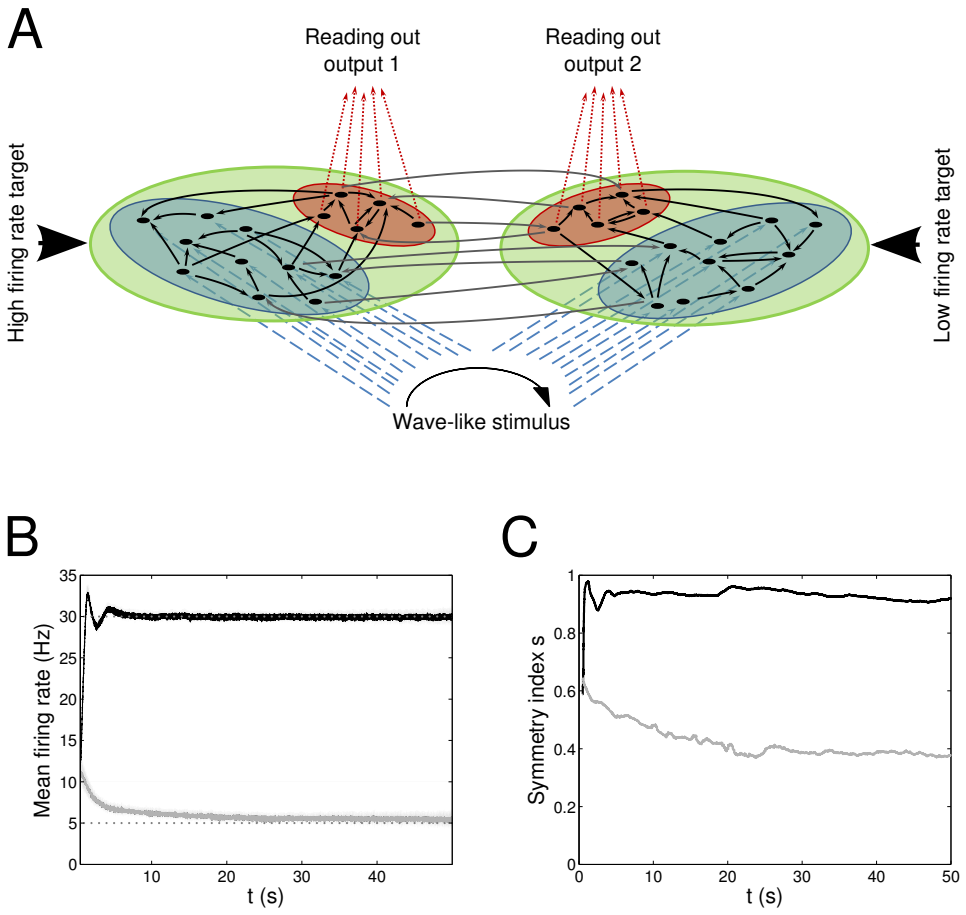

Figure 5.4 Double population scenario: Network architecture, activity and connectivity of the output populations with full (U, τ_{rec}, τ_{facil}, A) learning scheme (*Part 1*). **A** Architecture. The previous network is doubled so that there are now four populations: Two input regions (*blue*) and two output regions $\wp_1^{out}$, $\wp_2^{out}$ (*red*). The four populations are organised in two branches, one required to fire at high rates ($30\,Hz$) and the second at low rates ($5\,Hz$). Within each branch connections are all to all (*black arrows*) whereas initially weak connections (*grey arrows*) are present between the two output populations and between the two input populations. Input neurons receive a wave-like stimulus from outside (*blue dashed arrows*). All synapses obey both Spike-Timing Dependent Plasticity and rate-dependent short-term plasticity. **B** Mean firing rate of the output populations, *black line* for $\wp_1^{out}$ and *grey line* for $\wp_2^{out}$. *Shaded area* represents standard deviation and *horizontal dotted grey lines* show the two target firing rates ($30\,Hz$ for $\wp_1^{out}$, $5\,Hz$ for $\wp_2^{out}$). **C** Symmetry measure applied on the connectivity of the output population. Colour legend as in panel B. Connectivity evolves differently in the two populations, leading to a bidirectional motif in $\wp_1^{out}$ and to a unidirectional motif in $\wp_2^{out}$.

Network architecture

The new configuration is depicted in Fig. 5.4A and it is obtained by mirroring the structure of the first scenario and by adding recursive connections between functionally homologous populations. This led to a network of $N = 80$ conductance-based integrate-and-fire neurons,

organised in two distinct branches of 40 neurons each, with the first branch required to fire at a high rate ($\nu = 30Hz$) and the second branch at a low rate ($\nu = 5Hz$). Both targets remain fixed throughout the entire simulation. Each branch is a replication of the architecture we used previously, i.e. it is formed by an input and an output population recursively connected. Thus, the network is formed by four functionally different populations: $\wp_1^{in}$, $\wp_2^{in}$, $\wp_1^{out}$, $\wp_2^{out}$, with obvious meaning of symbols. Input populations in both branches receive the stimulus from the same source: A single wave-like signal is delivered to the $N_{input} = 60$ neurons with $\nu = 10Hz$, stimulating one neuron per time (see Methods 2.5), first the neurons in $\wp_1^{in}$ and then the neurons in $\wp_2^{in}$. All connections are plastic following the STDP triplet rule and TM model for STP.

Lateral connections are present between the inputs $\wp_1^{in}$, $\wp_2^{in}$ and between the outputs $\wp_1^{out}$, $\wp_2^{out}$. To stress that they are functionally different, we drew their initial values from a uniform distribution in $\left[10^{-3}, 10^{-1}\right]$, but, during the evolution, synapses are allowed to grow up to $A_{max} = 1$ as any other synapse. Furthermore, cross connections between different output and input populations, i.e. between $\wp_1^{in}$, $\wp_2^{out}$ and between $\wp_1^{out}$, $\wp_2^{in}$ are absent. The rest of the connections - within each population and across populations belonging to the same branch - are drawn from a uniform distribution in $\left[10^{-3}, 1\right]$ and they are not allowed to exceed this interval during the simulation. STP variables are initialised as in the single population scenario and in all the simulations presented in this subsection we used $\gamma = 2$ as the learning rate.

Full model: Adaptation of U, τ_{rec}, τ_{facil} and A

We begin by studying the behaviour of the *full model*: All four parameters are modified by our rate-dependent STP, Eq. (5.20)-(5.24). Taking into account the modifications of all three STP parameters allows us to make a direct comparison with Wang et al. [2006]. Results are displayed in Fig. 5.4B-C and in Fig. 5.5.

Fig. 5.4B,C shows the time course of the mean firing rate and symmetry index in both output populations, *black lines* for $\wp_1^{out}$ and *light grey lines* for $\wp_2^{out}$. *Shaded areas* and *dark grey dotted lines* represent standard deviation and target firing rates. Both populations $\wp_1^{out}$, $\wp_2^{out}$ approach the target rate while developing specific connectivity: As expected, a bidirectional motif emerges in the population firing at the high rate whereas the population firing at the low rate develops mostly unidirectional connections.

Fig. 5.5A-C shows the time evolution of the three parameters of the TM model: *Black lines* and *grey lines* represent the mean value of the synapses projecting from the two output populations $\wp_1^{out} \bigcup \wp_2^{out}$ respectively onto $\wp_1^{out}$ and $\wp_2^{out}$. Shaded area is the standard deviation. As expected from the previous simulation, we observe that the two populations

develop different synaptic types: High values of τ_{facil} and low values of τ_{rec} and U, as observed in the population firing at the high rate, suggest a facilitating behaviour, whereas values as the one observed in $\wp_2^{out}$, characterise depressing synapses. Mean values at the end of the simulation are reported in Tab. 5.3 *rows 1,4*. These results show that our model develops target-specific STP and results in good agreement with the data in Wang et al. [2006]. Indeed, although single values are not identical, the qualitative synaptic behaviour is represented: Recalling the notation used in Wang et al. [2006], two main types of synapses are present. The group projecting from $\wp_1^{out} \bigcup \wp_2^{out}$ onto $\wp_1^{out}$ can be mapped onto the type $E1$ and the group projecting from $\wp_1^{out} \bigcup \wp_2^{out}$ onto $\wp_2^{out}$ that can be mapped onto the type $E2$.

Following Wang et al. [2006], we can also refine our classification, introducing a further distinction within each class. With this purpose, we show in Fig. 5.5D-F the distributions of τ_{rec}, τ_{facil} and U at the end of the simulation within the entire output population $\wp_1^{out} \bigcup \wp_2^{out}$. For each histogram, data have been divided into four groups, representing the four different subtypes: $\wp_2^{out}$ to $\wp_2^{out}$ with *light grey*, $\wp_1^{out}$ to $\wp_2^{out}$ with *medium grey*, $\wp_1^{out}$ to $\wp_1^{out}$ with *dark grey*, $\wp_2^{out}$ to $\wp_1^{out}$ with *black*. While the distinction between the two synaptic types mapping onto $E1$ and $E2$ is evident, the difference between two subtypes in the same type cannot be easily seen. However, by looking at the mean values of synaptic parameters in Tab. 5.3 *rows 2, 3, 5, 6* and in particular the ratio τ_{rec}/τ_{facil} in Tab. 5.3 *column 5*, the distinction into four subtypes becomes more clear. As reported in *column 7* of Tab. 5.3, we can map the synaptic subtypes as follows: $E1a$ corresponds to the group $\wp_1^{out} \rightarrow \wp_1^{out}$, $E1b$ to $\wp_2^{out} \rightarrow \wp_1^{out}$, $E2a$ to $\wp_2^{out} \rightarrow \wp_2^{out}$ and $E2b$ to $\wp_1^{out} \rightarrow \wp_2^{out}$.

Finally, similarly to Fig. 5.2, in Fig. 5.5G-J we show single synapse traces for each subtype. We observe that, except for the last trace, different groups effectively show a distinctive response to the same stimulus ($12Hz$) and the traces reproduce the ones of the corresponding subtypes in Wang et al. [2006].

Although in Fig. 5.5D-F we present four different histograms for each parameter, we can reason on the overall distribution within the entire output population $\wp_1^{out} \bigcup \wp_2^{out}$ as the sample size is the same in all histograms. We can therefore observe that the distribution of τ_{rec} closely matches that in Wang et al. [2006], whereas the distribution of U reproduces the peak at around 0.25 but is less broad. The distribution of τ_{facil}, instead, is rather different, being totally shifted towards facilitating values in our case. This may be due to the fact that U is much more peaked around low values. We decided then to discard τ_{facil} from the learning scheme and run a simulation where only U, τ_{rec} and A are learnt, as we did for the single population scenario. We observed that the behaviour of the output populations and all the results remain unchanged. We provide an explanation for this in the Discussion.

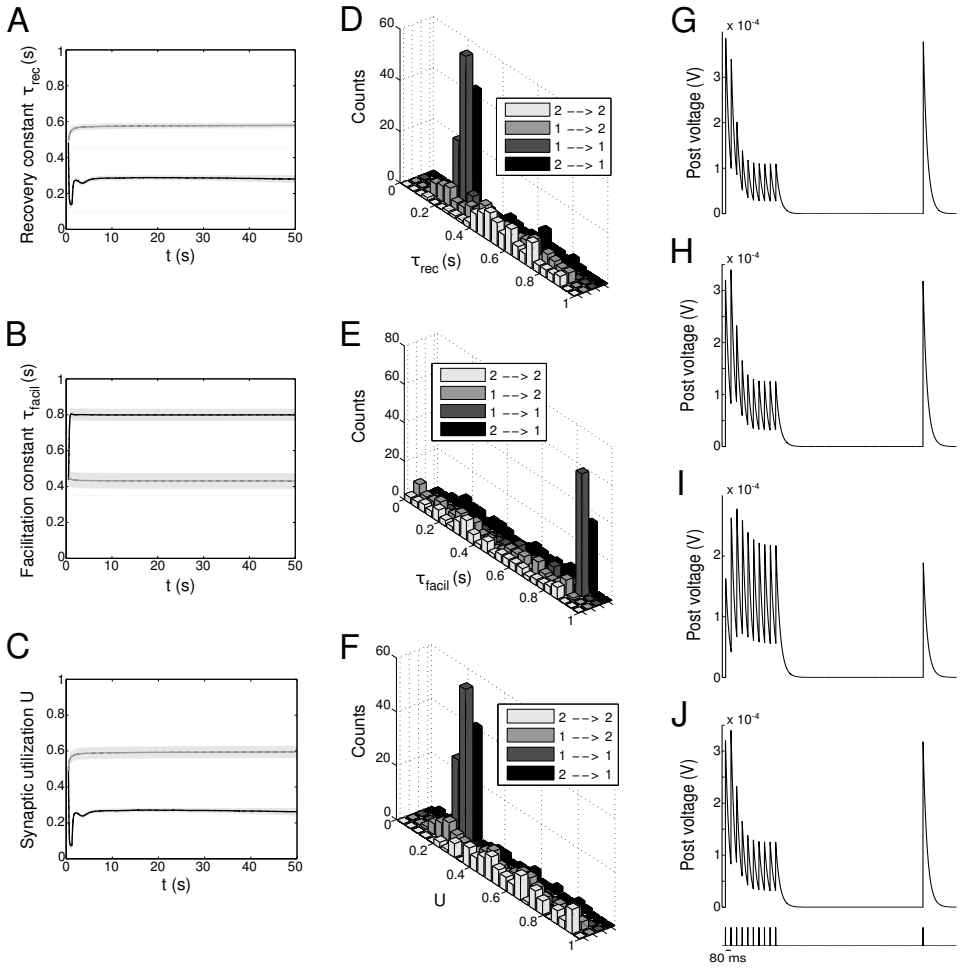

Figure 5.5 Double population scenario: STP parameters adaptation and final distribution for the output populations with full (U, τ_{rec}, τ_{facil}, A) learning scheme (*Part 1*). **A - C** Mean values of recovery time constant τ_{rec}, facilitation time constant τ_{facil} and synaptic utilisation U. *Black lines* represent mean values across the synapses projecting onto output population 1 from both output populations, $\wp_1^{out} \bigcup \wp_2^{out} \rightarrow \wp_1^{out}$, whereas *grey lines* describe the synapses projecting onto output population 2 from both output populations, $\wp_1^{out} \bigcup \wp_2^{out} \rightarrow \wp_2^{out}$. *Shaded areas* show standard deviation. We observe that the two populations develop different synaptic types, facilitating for $\wp_1^{out}$ and depressing for $\wp_2^{out}$. **D - F** Corresponding histograms of the three synaptic parameters at the end of the simulation. For each of them we show four different groups of values, mapping qualitatively to the four subtypes identified by Wang et al. [2006], see Tab. 5.3. *Light grey:* $\wp_2^{out} \rightarrow \wp_2^{out}$ (E2a). *Medium grey:* $\wp_1^{out} \rightarrow \wp_2^{out}$ (E2b). *Dark grey:* $\wp_1^{out} \rightarrow \wp_1^{out}$ (E1a). *Black:* $\wp_2^{out} \rightarrow \wp_1^{out}$ (E1b). **G - J** Single synapse traces obtained with the TM model by using a $12\,Hz$ stimulus. Each panel represents a different subtype of synapses. **G** $\wp_2^{out} \rightarrow \wp_2^{out}$. **H** $\wp_1^{out} \rightarrow \wp_2^{out}$. **I** $\wp_1^{out} \rightarrow \wp_1^{out}$. **J** $\wp_2^{out} \rightarrow \wp_1^{out}$. Synaptic parameters used are the mean values obtained from the distributions drawn in panels D - F. A comparison with Wang et al. [2006] on the basis of the traces only shows that we are able to identify three of the four subtypes.

Synaptic groups	$\tau_{rec}(ms)$
$\wp_1^{out} \bigcup \wp_2^{out} \rightarrow \wp_1^{out}$	300 ± 9
$\wp_1^{out} \rightarrow \wp_1^{out}$	267 ± 6
$\wp_2^{out} \rightarrow \wp_1^{out}$	327 ± 15
$\wp_2^{out} \bigcup \wp_1^{out} \rightarrow \wp_2^{out}$	524 ± 16
$\wp_1^{out} \rightarrow \wp_2^{out}$	486 ± 23
$\wp_2^{out} \rightarrow \wp_2^{out}$	567 ± 22

Table 5.4 Types and subtypes of excitatory synapses between the two output populations in the minimal model (τ_{rec}, A). Symbols are as in Tab. 5.3. Similar to Wang et al. [2006], we provide the results in the form $mean \pm s.m.e.$.

A minimal model for rate-dependent STP: Adaptation of τ_{rec} and A

Finally, we study the *minimal model*: A model that suffices to obtain the desired behaviours by adapting as few parameters as possible. The choice of the parameters to be learnt is naturally suggested by the form of the objective function Eq. (5.9): τ_{rec} and A. Interestingly, this minimal model preserves two key features: *i)* Both a presynaptic parameter, τ_{rec}, and a postsynaptic parameter, A, participate in learning. *ii)* STP and STDP are linked to each other through the STP-dependent modification of A.

In Fig. 5.6 we show the results of the minimal model: From A to D respectively: Mean output firing rates, symmetry index, τ_{rec} evolution and τ_{rec} distribution in the four groups of synapses. By comparing these panels with the ones from the full model simulation, we observe that output populations still efficiently fulfil the task while developing the expected connectivity motifs. Also, in Tab. 5.4 we report the mean values of τ_{rec} for the four groups of synapses that we identified with the full model: There is still a clear distinction between them. We can therefore conclude that this minimal model is sufficient for qualitatively reproducing the main two types and also the subtypes of Wang et al. [2006].

5.3 Discussion

It is well known that synapses are activity-dependent connections through which neurons propagate information. STP is a mechanism that describes these phenomena in short time scales and introduces two typical synaptic behaviours: Depression and facilitation. Contrary to long-lasting modifications of maximum synaptic strengths, for example STDP, existing models of STP do not rely on any learning mechanisms, apart from very few exceptions; see for instance Carvalho and Buonomano [2011]. Motivated by their work, it is our belief

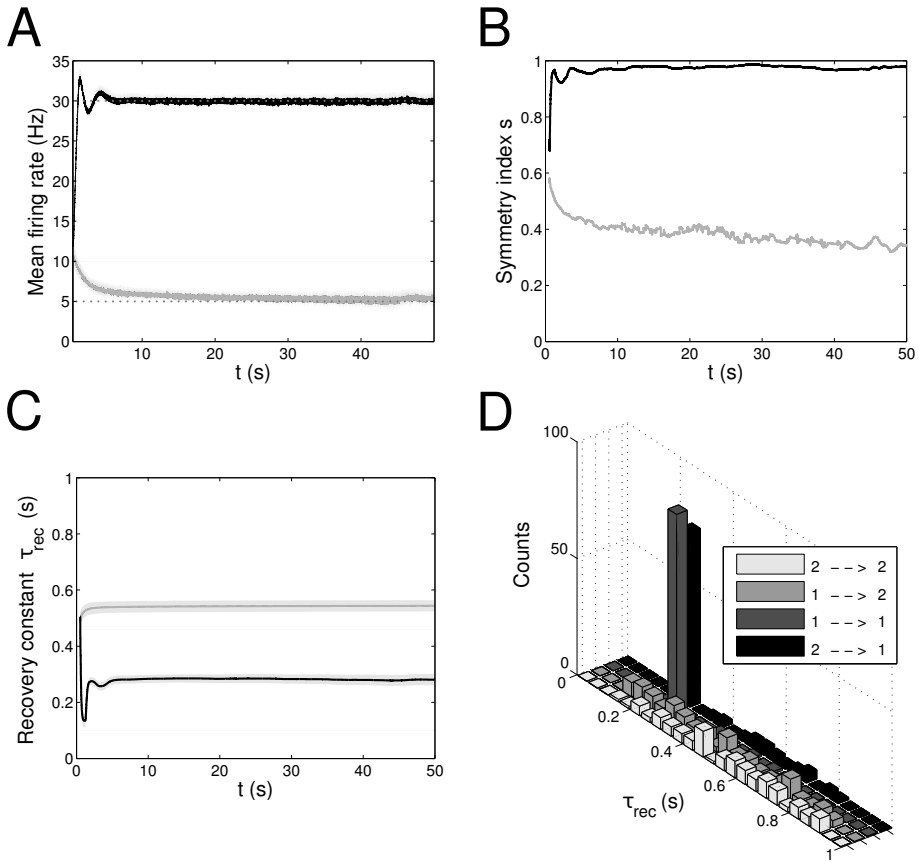

Figure 5.6 Double population scenario: Learning in the output populations with minimal (τ_{rec}, *A*) model. A Mean firing rate of the output populations, *black line* for $\wp_1^{out}$ and *grey line* for $\wp_2^{out}$. *Shaded area* represents standard deviation and *horizontal dotted grey lines* show the two target firing rates ($30\,Hz$ for $\wp_1^{out}$, $5\,Hz$ for $\wp_2^{out}$). **B** Symmetry measure applied on the connectivity of the output population. Colour legend as in panel B. Connectivity evolves differently in the two populations, leading to a bidirectional motif in $\wp_1^{out}$ and to a unidirectional motif in $\wp_2^{out}$. **C** Mean value of recovery time constant τ_{rec}. *Black line:* $\wp_1^{out} \bigcup \wp_2^{out} \rightarrow \wp_1^{out}$. *Grey line:* $\wp_1^{out} \bigcup \wp_2^{out} \rightarrow \wp_2^{out}$. We observe that the two populations develop different type of synapses, facilitating for $\wp_1^{out}$ and depressing for $\wp_2^{out}$. **D** Corresponding histograms of the recovery time constant at the end of the simulation. *Light grey*: $\wp_2^{out} \rightarrow \wp_2^{out}$, *medium grey*: $\wp_1^{out} \rightarrow \wp_2^{out}$, *dark grey*: $\wp_1^{out} \rightarrow \wp_1^{out}$, *black*: $\wp_2^{out} \rightarrow \wp_1^{out}$. The panels show that the achievement of the tasks and the differentiation of the synapses is still possible with this minimal model.

that more efficient dynamics would be possible if synapses were allowed to change their short-term behaviour by tuning their own parameters, depending on one or more external controlling factors, for example, their current task. Typically, one asks which is the firing regime for which a certain type of synapse performs better (Barak and Tsodyks [2007]), whereas we are looking at the picture from a reverse perspective: We want to obtain some frequency regime, which is the most efficient way to do it from a synaptic point of view?

A similar concept can be found in Natschläger et al. [2001], where the authors trained a network with a temporal structured target signal, using optimisation techniques.

In our work, we developed a learning scheme for STP, and we obtained, with a semi-rigorous argument, a learning rule for only one of the three parameters of the TM model, τ_{rec}. Based on specific experimental results (Markram et al. [1998b]; Thomson [2000]; Tsodyks and Markram [1997]) and data fitting (Chow et al. [2005]), we used the conjecture that STP behaviour of synapses has the same functional dependence on U and τ_{rec}, which allowed us to write a similar rule for the synaptic utilisation U. Interestingly, such learning rules depend on the maximum synaptic strength, and they therefore: *i)* Provided a natural link between STP and STDP and *ii)* allowed us to derive an STP-dependent rule for the maximum synaptic strength, to be added to the STDP contribution. It is important to highlight that, despite the heuristic derivation of the learning rules does not guarantee that they follow the gradient, our results show that the objective function Eq.(5.9) monotonically decreases with the adaptation of the parameters, eventually reaching its global minimum. This is a direct consequence of the fact that in each dynamical phase the average firing rate approaches the target firing rate, driving the value of objective function towards zero.

The interaction between short- and long-term plasticity is largely supported by experimental evidence (Markram et al. [1997]), although the exact mechanisms are still unknown. Some results (Markram and Tsodyks [1996]; Sjöström et al. [2003, 2007]) suggest that synapses become more/less depressing after long-term potentiation/depression. Our rules incorporate this behaviour: Long-term potentiation/depression always produces larger/smaller changes in STP parameters. However, whether these modifications bring more facilitation or depression critically depends on whether the population firing rate $\langle \nu \rangle$ is approaching the target rate ν_{targ} from above or below. Consider, for example, Eq. (5.16): If $\nu_{targ} - \langle \nu \rangle < 0$, then long-term potentiation will produce a stronger depression, thus reproducing the experimentally observed behaviour. In our simulations, this happens to the neurons that are firing at low frequencies. If $\nu_{targ} - \langle \nu \rangle > 0$, then an increase in A will make τ_{rec} even smaller, resulting in a less depressing synapse. In our simulations, this happens to the neurons that are firing at high frequencies. A similar argument can be formulated for the induction of long-term depression. We note that several mechanisms have been identified to compete during synaptic transmission, resulting in a more complex and less clear relationship between STP and STDP (Sjöström et al. [2007]).

In Sjöström et al. [2003, 2007] the authors link the interaction between short- and long-term plasticity with the frequency of firing: At high rates, synapses tend to become stronger and more depressing, while at lower frequencies they tend to become weaker and less depressing. Our derivation, instead, suggests the opposite: If we rely on the

hypothesis that large values of τ_{rec} lead to depression and small values to facilitation (Chow et al. [2005]), according to Eq. (5.15), facilitating synapses allow neurons to reach higher frequencies. These findings, together with the STDP triplet rule, form the basis of our work: They provide the theoretical basis for the experimentally observed correspondence between facilitation and bidirectionality, and between depression and unidirectionality. The behaviour expressed by Eq. (5.15) is experimentally and computationally based on previous work that relates facilitation with high frequency and rate code, and depression with low frequency and temporal code (Blackman et al. [2013]; Fuhrmann et al. [2002]). This is because, for example, a facilitating synapse may require several spikes to elicit an action potential, meaning that only high frequency stimulation can generate postsynaptic spikes (Klyachko and Stevens [2006]; Matveev and Wang [2000]).

We derived our rules by minimising an error function that is equal to zero when the target and actual firing rates are equal. Alternatively, we could have defined a reward function opposite to the error function in the sense that for zero error the reward function has its maximum value, and it is equal to zero for large error. We could have then taken the gradient of the reward function instead, bringing the derived rules into the framework of policy gradient learning methods and reinterpreting the feedback signal as a reward signal (Richmond et al. [2011]; Urbanczik and Senn [2009]; Vasilaki et al. [2009a]). In biological systems, dopamine is thought to act as reward signal (Fiorillo et al. [2003]; Schultz et al. [1997]), and its role in the context of learning associated with STDP, and more generally with Hebbian learning, has been extensively studied (Izhikevich [2007]; Legenstein et al. [2008]; Tobler et al. [2005]).

As highlighted in the introduction of this chapter, adaptation of STP parameters is a topic that has received very little attention, hence there is a lack of experimental results that could validate our hypothesis or provide some biological basis. The only available results in this direction, to the best of our knowledge, concern multiple evidence of activity-dependent vesicle replenishment, providing experimental basis for our adaptation of τ_{rec} (Hennig [2013]). Moreover, few models have been developed to describe this phenomenon, bringing with them possible biological interpretations. The most accepted one involves calcium concentration at the presynaptic terminal (Fuhrmann et al. [2004]). Given this evidence and the predominant role of calcium in both facilitation and depression, adaptation of STP parameters could be linked to a retrograde signal triggered by the postsynaptic neuron which would regulate the level of Ca^{++} in the presynaptic cell. The existence of a retrograde comunication between post and presynaptic neurons has been validated by experimental observations (Bender et al. [2006]; Sjöström et al. [2003, 2007]) and its specific role in interacting with calcium concentration to enhance facilitation or depression has been also

suggested by Carvalho and Buonomano [2011]) to support their model of STP adaptation. Due to the form of our rules, also the presence of two global feedback signals is required: one encoding the population activity, which is processed outside the population and broadcasted to all neurons; and an external signal bringing information about the current paradigm, i.e. the target firing rate. Similar to Urbanczik and Senn [2009], we can assume that synapses receive both signals via ambient neurotransmitter concentrations, leading to an on-line plasticity rule.

Finally, our derivation of the learning framework leads to a rather unsual STP-dependent contribution to the modification of the synaptic weight A. Its structure is similar to the learning rules for STP parameters and it has to be interpreted as a contribution that sums up with the traditional long-term plasticity (STDP in our case). This term is of difficult interpretation and there is no other mechanism such this in the literature. Therefore, possible biolgical interpretations have to remain within the domain of pure speculation. Receptor desensitisation (Hennig [2013]), for instance, is a property that has been shown to contribute to synaptic depression through a decrease in the number of available receptors on the postsynaptic site, due to presynaptic activity. As such, it represents an instance of modification of a postsynaptic quantity due to presynaptic activity, that is similar to what happens in our learning rule for A_{ij}^{STP} Eq. (5.23).

We initially tested our learning scheme by implementing the rules for τ_{rec} and U on a classical paradigm of inverting associations: Keeping the stimulus fixed and varying the associations, the network had to learn to first make choice A and then unlearn it in favour of choice B. This led to a network able to periodically switch its behaviour from depressing to facilitating and vice versa, closely following the change in the association paradigm. Throughout the simulation, the network formed motifs similar to those experimentally observed in Pignatelli [2009]; Wang et al. [2006], with facilitating synapses developing bidirectional motifs and depressing synapses developing unidirectional motifs. The desirable motifs were formed due to two factors: (i) The triplet rule that governed long-term potentiation and (ii) the wave-like input stimulus of the network. The form of the plasticity rule guarantees that when neurons fire at high frequency, the synaptic efficacy increases. Hence, synapses will grow up to their bounds, leading to bidirectional connections. On the contrary, when neurons fire at low frequencies, the synaptic efficacy decreases, yet the wave-like input imposes unidirectional connectivity.

We further extended this learning model by adding an STP-motivated rule for the maximum synaptic strength, and we tested it on the same invert association scenario. Results showed the same behaviour as before but with faster dynamics due to the joint action of STP and STDP on the absolute efficacy.

In the second part of the paper, we extended our study. First, we considered two populations that have to fire at different frequencies (low, high). Then, we introduced a learning rule for the facilitating time constant, in order to have a full learning model involving all four parameters. The aim was twofold:

(i) Comparison of our results with experimental data in Wang et al. [2006]. Although the accuracy is not excellent, we were able to qualitatively reproduce the basic differentiation in the ranges of values of the STP parameters, reflecting the existence of four different synaptic subtypes. We believe that by further adapting the model, in particular learning rates and target frequencies or by considering other rule combinations, it is possible to obtain different parameter values (in principle an infinite combination of them), and thus possibly reproduce the results of Wang and collaborators even better. However, we think this may not be critical because, as a recent study (Costa et al. [2013]) has pointed out, fitting techniques generally used for deriving STP parameters from experimental data may give unreliable results. Given this limitation, we think it is important that our model accounts for a large variety of parameter values in principle, and that in this specific case of Wang et al. [2006] it is able to replicate the basic distinction in the synaptic response.

(ii) Differentiation of synaptic types innervating two functionally different populations. The reason for this lies in the way we constructed the learning model: What triggers the synaptic modification is the spike of the postsynaptic neuron. The firing rate of the population to which this postsynaptic neuron belongs is the information used to tune the values of STP parameters. In other words, we implement a target-specific learning mechanism. This choice is based on an optimisation argument: The more direct and efficient way for a neuron to influence its own activity through synaptic changes is to modify incoming synapses rather than outgoing synapses. A second scheme, a source-specific learning mechanism modifying the outgoing synapses, would have probably led to the same results within closed microcircuits, but on a much longer time scale.

Our target-specific learning mechanism is also supported by experimental evidence (see Blackman et al. [2013] for a review). Despite the fact that STP seems to be mainly a presynaptic mechanism, it has been shown that the target cells can also determine the STP dynamics. All the studies we are aware of have established such a target specificity only in the context of excitatory cells innervating other excitatory cells on one side and inhibitory cells on the other side, specially interneurons (Buchanan et al. [2012]; Markram et al. [1998b]; Reyes et al. [1998]). It would therefore be interesting to appropriately modify the double population scenario by incorporating a population of inhibitory neurons and comparing the results with existing data. In addition, some authors (Blackman et al. [2013]; Costa et al. [2013]) suggested that a similar differentiation might exist within excitatory only populations.

Having target-specific STP for excitatory-excitatory connections is still an open possibility that needs to be further explored. Here we show from a theoretical point of view that such a differentiation is possible between fundamentally similar (all excitatory) but functionally different (encoding for different paradigms) targets.

The well-established existence of STP-target specificity provides us with a possible biological explanation for the learning rules we derived. Indeed, this scenario requires that the postsynaptic neuron can regulate specifically its own presynaptic compartment only, by a retrograde signal that does not affect neighbouring cells. Thus, diffusive retrograde messengers, for example endocannabinoids and nitric oxide, do not appear to be the most suited agents, whereas synaptic adhesion molecules, for example cadherins (Bozdagi et al. [2004]) and neuroligins (Dean and Dresbach [2006]), seem to be better candidates for playing this role. These molecules are responsible for governing the presynaptic transmitter release through many different presynaptic mechanisms (Blackman et al. [2013]; Blatow et al. [2003]; Deng et al. [2011]; Zucker and Regehr [2002]).

We underline that the way we obtained the learning rules is based in part on heuristic evaluation. According to Eq. (5.15), derived from a semi-rigorous argument, the key parameters seems to be τ_{rec} and A. By also including U following Carvalho and Buonomano [2011], we obtain a learning scheme involving τ_{rec}, U and A only, which we used to study the double population problem and evaluate the importance of τ_{facil}. Results remain essentially unchanged from the full model, suggesting that τ_{facil} does not play a critical role in the task we defined. This is not surprising and the reason is that our rules link facilitation with a high firing rate, and depression with a low firing rate. Indeed, even with a small facilitation time constant (small τ_{facil}), synapses are still able to fire at a high rate, as long as the stimulating frequency is high enough and recovery from depression is fast enough (low τ_{rec}). Therefore, the time constant of recovery from depression seems to be the only parameter regulating the firing frequency of the neuron for high firing rates, exactly as it comes out from the objective function (we recall that Eq. (5.15) comes from an inequality obtained in the limit of high frequency). With our novel view of allowing synapses to modify their properties from facilitating to depressing and vice versa, we therefore suggest that τ_{rec} is the parameter that is mostly related to rate coding, whereas U to temporal coding.

This conclusion is also supported by Carvalho and Buonomano [2011]. In this paper the authors described a simple problem based on temporal synchrony between two inputs that cannot be solved unless STP is learnt, together with STDP. Besides the long-lasting change in A, they introduce a temporal synaptic plasticity for U only and they showed that this indeed solves the problem. Also, they reported that changing U only was the most efficient way to solve the problem. Our work supports the hypothesis that, when dealing with

rate coding tasks, the only necessary parameter that has to be learnt is τ_{rec}, whereas, based on Carvalho and Buonomano [2011], when dealing with temporal coding tasks, the only necessary parameter is U.

Another result pointing to a similar direction can be found in Natschläger et al. [2001], where the authors use optimisation techniques, rather than explicit learning rules, to train a network of neurons in order to transform a time-varying input into a desired time-varying output. They show that to achieve good performance, one needs to change at least two parameters, either A and τ_{rec}, or A and U. This confirms that learning must involve at least one presynaptic and one postsynaptic parameter, and that τ_{facil} seems not to be relevant in these types of tasks.

We finally presented results from what we call the minimal model, where only τ_{rec} and A were allowed to change, since both their corresponding update rules come directly from the gradient of the objective function we defined. Results confirmed our belief, as we were still able to learn the tasks while obtaining results similar to those from Wang and collaborators. It is in agreement with our conjecture that when we tried to apply learning on U and A only (results now shown here), the network failed to perform its task because the population that was supposed to fire high stabilised at a much lower frequency, i.e. $\sim 15\,Hz$. Therefore, an alternative minimal model adapting U and A would be able to successfully learn only targets of a lower firing regime. We believe that specialisation of parameters in the STP model depending on tasks and signal encoding may be a key ingredient towards a better understanding of synaptic and neuron functionality.

Summary The mechanism described here suggests a possible answer to the question why we observe a correspondence between certain kinds of motifs and short term properties of synapses. It is innovative not just for the form of the rule, but also for the rule itself: STP is usually thought as a phenomenon that depends on long-term plasticity, without a separate and independent learning mechanisms. The model presented is a simplistic one that wants to show how learning STP is possible and it reproduces experimental results, based on the idea that it is "convenient" for both synapses and neurons, and therefore for the computation in general. In the next Chapter possible improvements for this study will be discussed, as well as for the rest of the work presented in Chapters 3 and 4.

Chapter 6

Future Developments

The way neurons are connected within the brain is not random, but follows specific patterns. Even though brain connectivity follows a general scheme that is recurrent among humans, each individual develops his own wiring diagram that is different from anyone else, as it is shaped by genes and experience, and therefore also by possible disorders. The map of brain connections is thought to be the key to access knowledge that can shed new light on a variety of brain mechanisms, in terms of single processes and brain functioning as a whole. Revealing connectivity patterns is therefore essential. Being able to interpret these patterns, or motifs, and to identify the causes that made them develop and have the observed features is just as important. Considerable work has been done in this direction and more is still ongoing.

In this thesis, these two big challenges in Neuroscience were formulated in terms of *how* and *why* questions: How can we reveal motifs in large brain-like networks? And, can we give an explanation of why such motifs seem to appear correlated with specific synaptic properties? The work presented here attempts to address both questions. First, we formulated the problem of finding and evaluating structures in networks of neurons in terms of a community detection problem. We therefore introduced a symmetry measure definition and we constructed a statistics-based heuristic algorithm for directed and weighted networks aiming at identifying overlapping bidirectional communities in large networks. On a separate, but strictly related line of research, we also introduced an error-driven learning mechanism for short-term plasticity that may explain how such motifs, or communities, develop and correlate with synaptic properties, starting from random initial conditions. In particular, we were able to qualitatively reproduce the experimental results of Wang et al. [2006]. The work presented in this thesis is mainly taken from the papers I produced: Esposito et al. [2014, 2015]; Esposito and Vasilaki [2015].

The results showed in this thesis have, however, their own limitations and also they cannot account for all the aspects of the connectivity problem. As such, they do not represent an end point, rather they should be regarded as a source of inspiration and as a staring point for improvements and further studies. Some of the possible ideas have been already mentioned at end of each chapter and will be further discussed below.

Motifs detection

It is important to stress that the symmetry measure and the algorithm for motifs detection have been tested on *in silico* data, and indeed they are mainly intended to be a tool for the computational neuroscience community. This is mostly due to the fact that experimental recordings do not provide connectivity matrices but sequences of spikes. However, ongoing efforts of inferring synaptic weights and reconstructing the effective connectivity from spikes suggest that in a near future the tools provided here could be applied also to experimental data (Friston [2011]; Pernice and Rotter [2013]; Van Bussel et al. [2011]).

Connectivity prior to learning A major limitation in our procedure for communities detection is that it requires the knowledge of the distribution of synaptic weights prior to learning. When simulating the behaviour of a network of neurons, in most cases such a distribution is imposed at the beginning, and it is therefore known. Thus, the relevant question becomes how to choose such a distribution. Here, as a case study, we considered uniform and Gaussian distributions, but in principle all the results can be reproduced for a different distribution. The only requirement is that the distribution needs to be regular, i.e. one for which it is meaningful to apply measures from statistics, like mean value and variance.

In terms of experimental data, the appropriate question would be rephrased as: How can we know which is the real distribution of connections in a brain circuit before learning takes place? A possible answer is certainly related to neural development, the study of the brain during early stages of development. This topic has been studied for years, with different methods suggested to analyse cortical development and its implications (Kolb and Gibb [2011]; Mariani et al. [2012]; Mills and Tamnes [2014]; Muzio and Consalez [2013]). Recently, due to the increasing interest towards the connectome, there has been a wealth of research looking at how the connectome itself changes over time (Ingalhalikar et al. [2015]; Mills and Tamnes [2014]). Remarkably, a recently launched program, called Developing Human Connectome Project[1], aims to create the first connectome of early life. Of course, even during development, there is no clear distinction between a hypothetical pre-learning

[1] http://www.developingconnectome.org/

phase and what happens afterwards. It is probably not possible to obtain such an entire map without including effects due to learning processes: As soon as connections form, they are extremely plastic and therefore they are already subject to modifications. Despite such complications, having access to the history of connectivity, from the very beginning of development to later stages, would certainly be helpful in clarifying how connections between neurons could be distributed. Also numerical simulations would benefit from this as they could rely on a more realistic basis.

Sparseness A second key aspect is that connectivity within brain circuits is sparse, whereas the networks we considered to test our algorithm are all-to-all connected, and the algorithm itself is built for this types of graphs. The algorithm presented in Chapter 5 is our first attempt to address the structure detection problem, and as such we chose a configuration as simple as possible, i.e. with neurons all-to-all connected. The good results obtained encourage us to develop it further for more realistic and complicated cases, such as including sparseness.

We recall that the algorithm relies on two elements: The symmetry measure, which already takes sparseness into account through the pruning parameter a, and the bidirectional community definition, that was conceived for all-to-all connected networks. The way we built such a definition allows a very immediate generalisation to sparse bidirectional communities: Similar to what we did for the symmetry measure, we can introduce a pruning parameter a and rescale the minimum number of bidirectional connections required within the community. Thus, similar to Eq. (4.5) and with the same meaning of symbols, we can define sparse bidirectional communities as a set of neurons $\mathscr{C}$ such that:

$$|\mathscr{S}^i_{\mathscr{C}}| \geq a\vartheta_{\mathscr{C}}\left(|\mathscr{C}|-1\right) \quad \forall i \in \mathscr{C}\,. \tag{6.1}$$

A further development would be to better characterise sparseness, for example by using second or higher order statistics. Indeed, brain networks show a variety in the sparseness of the connectivity that is richer than what we can obtain with a single parameter, that results in a uniformly pruned network, hence globally regular. For instance, this would allow to take into account that, even though the global connectivity is sparse, locally there might be a higher probability of connection just because of anatomical reasons, leading to the formation of small local clusters. Indeed, significant presence of higher order motifs has been confirmed by experimental data (Song et al. [2005]; Sporns et al. [2005]) and simulations (Bourjaily and Miller [2011a]).

Small-world and scale-free topologies Recently, it has been proposed that brain networks may exhibit small-world (Bassett and Bullmore [2009]; Sporns et al. [2004]; Yu et al.

[2008]) and scale-free properties (Beggs and Plenz [2003]; El Boustani et al. [2009]; He [2014]; Levina et al. [2009]). These results are extremely attractive because these types of organisation seem to better respond to optimality criteria in terms of information transmission. Also, what makes these graphs highly interesting is that they have been found in many different systems in nature and they are related to the concept of local cluster. A natural development of the work presented here would therefore be the adaptation of our algorithm in order to apply it to networks with small-world and scale-free properties.

For example, in the Watts-Strogatz model (Watts and Strogatz [1998]) the community structure intuitively depends on the rewiring probability p: The extreme value $p = 1$ generates a random graph, which has no community structure, while $p = 0$ generates a regular lattice. The community detection algorithm could therefore be used, for instance, to establish the transition value of the parameter p.

In scale-free networks, by definition, nodes with fewer connections are the candidates to form communities, resembling a small-world structure. Most of the scale-free networks, indeed, show the property of small-worldness. However, this is not always the case: The Barabási-Albert model (Albert and Barabási [2002]), for instance, is a scale-free network that does seem to have a community structure (Fortunato [2010]), because the connection between each pair of nodes is random due to the preferential attachment procedure. An existing tool to evaluate the clustering properties of a network is the clustering coefficient (Watts and Strogatz [1998]): Given 3 neurons i, j, k, if both j and k are connected to i then the probability that there is also a connection between j and k is high (higher than in random graphs). Intuitively, the clustering coefficient appears to be a necessary but not sufficient condition for community structure, as the definition relies on motifs of 3 neurons and therefore it does not allow to infer much about larger motifs. Indeed, even though not all scale-free networks exhibit a community structure, both small-world and scale-free networks are characterised by a high value of the clustering coefficient. Our algorithm could therefore be used to better distinguish these two classes of graphs and also to separate within each class networks having a community structure from those that do not, giving further insights on their topology.

In order to apply our algorithm to small-world and scale-free networks some adaptations are required, as when dealing within small-world and scale-free networks, it is implicit to refer to binary connections: Small-world and scale-free, and therefore community structure, are features that are measured based on the presence/absence of a connection. Hence there is no involvement of connection weights and therefore no notion of bidirectionality or similar. This means that our algorithm cannot be directly applied to such networks, but modifications

are needed to adapt it to the context. Here we lay the basis for future study by speculating on two possible procedures:

- Prior to learning, a binary connectivity can be generated by following the desired small-world or scale-free network model; links that have been generated can be turned into continuous connections according with some distribution of weights; finally, a learning process can be applied. In this sense, the small-worldness would be just a different way of applying sparseness.
- Another possible way would be to generate the all-to-all network with continuous weights, as we did in our study, and afterwards, but prior to learning, binarise it.

We aim at investigating these possibilities so as to adapt and generalise our algorithm to different scenarios.

Benchmarks As mentioned above, the Barabási-Albert network is an instance of scale-free network with no community structure in it. This network can therefore be used as a first test for every algorithm aiming at detecting communitites. Considering the issues outlined above, we adapted our algorithm to be able to perform a community search in an Albert-Barabasi network and, correctly, no community structure was detected. To further test the quality of the performance of the algorithm we aim at testing it on a class of benchmark graphs that have been proposed as a measure to compare different algorithms (Lancichinetti and Fortunato [2009]).

Learning mechanisms

The mechanism we introduced in Chapter 5 is defined within a simple scenario, where, given a periodic input signal that is always the same, the desired output changes, forcing synapses to change as well. It should be stressed that this model is not intended to provide a realistic biological mechanism, but rather to introduce the concept of learning STP, to show how this can be done and how it has the potential of reproducing some experimental results. This can be used both as a baseline for more complex and interesting scenarios that can be defined to test the behaviour of the learning framework and for adjustments of the rule itself and alternative formulations.

Reinforcement Learning and segregation of synapses To make our rule more biologically plausible, it is our aim to rephrase it within the context of Reinforcement Learning. The idea is that different synapses can learn to selectively respond to different input patterns so

that the entire network is able to respond differently to different signals. In other words, the aim is to reproduce associative learning paradigms, that can be already obtained for instance with a perceptron, but with the novelty of a short-term learning rule and with a biologically plausible global signal that delivers reward or punishment to all the neurons in the network.

This different formulation reflects an alternative optimality criterion at the basis of the learning rule: With the mechanism introduced in Chapter 5, all neurons can change their synapses in order to reach the desired goal. With the new reinforcement learning framework, instead, only some neurons will be recruited, according to the needs: After a training phase, where changes in short-term properties of synapses are modulated by reward, when presenting the network with a specific input, only those synapses that are best suited for achieving the corresponding desired output will be activated for the transmission of the signal itself.

Is STDP top-down? The learning framework we defined, comprising both STP and STDP, results from a mixture of two different approaches. The learning rule for STP, which turns out to influence also the long-term plasticity (see Chapter 5), is the result of a gradient descent procedure: We defined an objective function and we assume that the system learns in order to minimise this global function. In other words, all the components of the system – the synapses in this case – have a global knowledge of it. This is called top-down approach (Crespi et al. [2008]). STDP, instead, is thought to be a bottom-up technique, as there is no global function to optimise: The learning rule comes directly from observation and the global behaviour emerges naturally as a consequence of the single local processes at synaptic level. Although using both approaches might seem contrasting, analysis of experimental data and results from modelling brain circuits show that bottom-up and top-down mechanisms can work together on the same process (Grossberg and Versace [2008]; Mechelli et al. [2004]; Szatmáry et al. [2013]). In particular, evidence has been found for bottom-up and top-down plasticity in orientation networks of the adult primary visual cortex V1 (Schummers et al. [2005]). These results open up the possibility that STDP also responds to a global optimisation criterion. Hence, we aim at investigating such a possibility through a mathematical analysis of the equations governing it.

Umberto Esposito
Sheffield, October 2015

References

Abbott, L. and Gerstner, W. (2004). Homeostasis and learning through spike-timing dependent plasticity. *Methods and models in neurophysics*.

Abbott, L. and Nelson, S. (2000). Synaptic plasticity: Taming the beast. *Nat Neurosci*, 3:1178–83.

Abbott, L. F. and Regehr, W. G. (2004). Synaptic computation. *Nature*, 431(7010):796–803.

Albert, R. and Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74.

Asaad, W. F., Rainer, G., and Miller, E. K. (1998). Neural activity in the primate prefrontal cortex during associative learning. *Neuron*, 21(6):1399–1407.

Babadi, B. and Abbott, L. F. (2010). Intrinsic stability of temporally shifted spike-timing dependent plasticity. *PLoS Comput Biol*, 6(11):e1000961.

Babadi, B. and Abbott, L. F. (2013). Pairwise analysis can account for network structures arising from spike-timing dependent plasticity. *PLoS Comput Biol*, 9(2):e1002906.

Bain, A. (1873). *Mind and body. The theories of their relation*. New York : D. Appleton and company.

Ball, B., Karrer, B., and Newman, M. E. J. (2011). Efficient and principled method for detecting communities in networks. *Phys Rev E Stat Nonlin Soft Matter Phys*, 84(3 Pt 2):036103.

Barabási, A.-L. and Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nat Rev Genet*, 5(2):101–113.

Barak, O. and Tsodyks, M. (2007). Persistent activity in neural networks with dynamic synapses. *PLoS Comput Biol*, 3(2):e35.

Bassett, D. S. and Bullmore, E. T. (2009). Human brain networks in health and disease. *Curr Opin Neurol*, 22(4):340–347.

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

Bender, V. A., Bender, K. J., Brasier, D. J., and Feldman, D. E. (2006). Two coincidence detectors for spike timing-dependent plasticity in somatosensory cortex. *J Neurosci*, 26(16):4166–4177.

Berlucchi, G. and Buchtel, H. A. (2009). Neuronal plasticity: historical roots and evolution of meaning. *Exp Brain Res*, 192(3):307–319.

Bi, G. and Poo, M. (1998). Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type. *J Neurosci*, 18(24):10464–72.

Bienenstock, E. L., Cooper, L. N., and Munro, P. W. (1982). Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *J Neurosci*, 2(1):32–48.

Billings, G. and van Rossum, M. C. W. (2009). Memory retention and spike-timing-dependent plasticity. *J Neurophysiol*, 101(6):2775–2788.

Binzegger, T. and Douglas, RJ Martin, K. (2004). A quantitative map of the circuit of cat primary visual cortex. *J Neurosci*, 24(39):8441–53.

Blackman, A. V., Abrahamsson, T., Costa, R. P., Lalanne, T., and Sjöström, P. J. (2013). Target-cell-specific short-term plasticity in local circuits. *Front Synaptic Neurosci*, 5:11.

Blatow, M., Caputi, A., Burnashev, N., Monyer, H., and Rozov, A. (2003). Ca2+ buffer saturation underlies paired pulse facilitation in calbindin-d28k-containing terminals. *Neuron*, 38(1):79–88.

Bliss, T. V. and Collingridge, G. L. (1993). A synaptic model of memory: long-term potentiation in the hippocampus. *Nature*, 361(6407):31–39.

Bliss, T. V. and Lømo, T. (1970). Plasticity in a monosynaptic cortical pathway. *J Physiol*, 207(2):61P.

Bliss, T. V. and Lømo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *J Physiol*, 232(2):331–356.

Bock, D. D., Lee, W.-C. A., Kerlin, A. M., Andermann, M. L., Hood, G., Wetzel, A. W., Yurgenson, S., Soucy, E. R., Kim, H. S., and Reid, R. C. (2011). Network anatomy and in vivo physiology of visual cortical neurons. *Nature*, 471(7337):177–182.

Bomze, I. M., Budinich, M., Pardalos, P., and Pelillo, M. (1999). *Handbook of Combinatorial Optimization*. Kluwer Academic Publishers, Norwell, USA.

Bondy, A. and Murty, U. (2008). *Graph Theory*. Springer.

Bourjaily, M. and Miller, P. (2011a). Excitatory, inhibitory, and structural plasticity produce correlated connectivity in random networks trained to solve paired-stimulus tasks. *Front Comput Neurosci*, 5(37):1–24.

Bourjaily, M. A. and Miller, P. (2011b). Synaptic plasticity and connectivity requirements to produce stimulus-pair specific responses in recurrent networks of spiking neurons. *PLoS Comput Biol*, 7(2):e1001091.

Bozdagi, O., Valcin, M., Poskanzer, K., Tanaka, H., and Benson, D. L. (2004). Temporally distinct demands for classic cadherins in synapse formation and maturation. *Mol Cell Neurosci*, 27(4):509–521.

Brandes, U., Delling, D., Gaertler, M., Görke, R., Hoefer, M., Nikolski, Z., and Wagner, D. (2006). *On Modularity - NP-completeness and Beyond*. Universität Karlsruhe, Fakultät für Informatik, Bibliothek.

Bressler, S. L. and Menon, V. (2010). Large-scale brain networks in cognition: emerging methods and principles. *Trends Cogn Sci*, 14(6):277–290.

Briggman, K. L. and Denk, W. (2006). Towards neural circuit reconstruction with volume electron microscopy techniques. *Curr Opin Neurobiol*, 16(5):562–570.

Briggman, K. L., Helmstaedter, M., and Denk, W. (2011). Wiring specificity in the direction-selectivity circuit of the retina. *Nature*, 471(7337):183–188.

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

Buchanan, K. A., Blackman, A. V., Moreau, A. W., Elgar, D., Costa, R. P., Lalanne, T., Jones, A. A. T., Oyrer, J., and Sjöström, P. J. (2012). Target-specific expression of presynaptic nmda receptors in neocortical microcircuits. *Neuron*, 75(3):451–466.

Bullmore, E. and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci*, 10(3):186–198.

Buonomano, D. V. and Merzenich, M. (1998). Cortical plasticity: From synapses to maps. *Annu. Rev. Neurosci.*, 21:149–86.

Carvalho, T. and Buonomano, D. V. (2011). A novel learning rule for long-term plasticity of short-term synaptic plasticity enhances temporal processing. *Front Integr Neurosci*, 5(20):1–11.

Cheetham, C. E. J., Hammond, M. S. L., Edwards, C. E. J., and Finnerty, G. T. (2007). Sensory experience alters cortical connectivity and synaptic function site specifically. *J Neurosci*, 27(13):3456–3465.

Chklovskii, D., Mel, B., and Svoboda, K. (2004). Cortical rewiring and information storage. *Nature*, 431(7010):782–8.

Chklovskii, D., Vitaladevuni, S., and Scheffer, L. (2010). Semi-automated reconstruction of neural circuits using electron microscopy. *Current Opinion in Neurobiology*, 20(5):667–75.

Chow, C., Gutkin, B., Hansel, D., Meunier, C., and Dalibard, J., editors (2005). *Methods and models in neurophysics*, chapter Synaptic Dynamics, pages 245–66. Elsevier.

Clopath, C., Buesing, L., Vasilaki, E., and Gerstner, W. (2010). Connectivity reflects coding: A model of voltage-based stdp with homeostasis. *Nat Neurosci*, 13:344–52.

Clopath, C., Ziegler, L., Vasilaki, E., Buesing, L., and Gerstner, W. (2008). Tag-trigger-consolidation: A model of early and late long-term-potentiation and depression. *PLoS Comput Biol*, 4(12):e1000248.

Cook, S. A. (1971). The complexity of theorem-proving procedures. In *Proceedings of the Third Annual ACM Symposium on Theory of Computing*, STOC '71, pages 151–158, New York, NY, USA. ACM.

Costa, R. P., Sjöström, P. J., and van Rossum, M. C. W. (2013). Probabilistic inference of short-term synaptic plasticity in neocortical microcircuits. *Front Comput Neurosci*, 7:75.

Crespi, V., Galstyan, A., and Lerman, K. (2008). Top-down vs bottom-up methodologies in multi-agent system design. *Journal of Autonomous Robots*, 24:303–313.

Danon, L., Duch, J., Diaz-Guilera, A., and Arenas, A. (2005). Comparing community structure identification. *J. Stat. Mech.*

Dayan, P. and Abbott, L. (2001). *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. The MIT Press: Cambridge, Massachusetts.

Dean, C. and Dresbach, T. (2006). Neuroligins and neurexins: linking cell adhesion, synapse formation and cognitive function. *Trends Neurosci*, 29(1):21–29.

Del Giudice, P., Fusi, S., and Mattia, M. (2003). Modelling the formation of working memory with networks of integrate-and-fire neurons connected by plastic synapses. *J Physiol (Paris)*, 97:659–681.

Deng, L., Kaeser, P. S., Xu, W., and Südhof, T. C. (2011). Rim proteins activate vesicle priming by reversing autoinhibitory homodimerization of munc13. *Neuron*, 69(2):317–331.

Denk, W. and Horstmann, H. (2004). Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. *PLoS Biol*, 2(11):e329.

Diestel, R. (2010). *Graph Theory*. Springer-Verlag Berlin Heidelberg.

Dittman, J. S. and Regehr, W. G. (1998). Calcium dependence and recovery kinetics of presynaptic depression at the climbing fiber to purkinje cell synapse. *J Neurosci*, 18(16):6147–6162.

Douglas, R. and Martin, K. (2007a). Mapping the matrix: the ways of neocortex. *Neuron*, 56(2):226–38.

Douglas, R. and Martin, K. (2007b). Recurrent neuronal circuits in the neocortex. *Curr Biology*, 17(13):R496–500.

Douglas, R. M. and Goddard, G. V. (1975). Long-term potentiation of the perforant path-granule cell synapse in the rat hippocampus. *Brain Res.*, 86:205–215.

Dunwiddie, T. and Lynch, G. (1978). Long-term potentiation and depression of synaptic responses in the rat hippocampus: localization and frequency dependency. *J Physiol*, 276:353–367.

El Boustani, S., Marre, O., Béhuret, S., Baudot, P., Yger, P., Bal, T., Destexhe, A., and Frégnac, Y. (2009). Network-state modulation of power-law frequency-scaling in visual cortical neurons. *PLoS Comput Biol*, 5(9):e1000519.

Esposito, U., Giugliano, M., van Rossum, M., and Vasilaki, E. (2014). Measuring symmetry, asymmetry and randomness in neural network connectivity. *PLoS One*, 9(7):e100805.

Esposito, U., Giugliano, M., and Vasilaki, E. (2015). Adaptation of short-term plasticity parameters via error-driven learning may explain the correlation between activity-dependent synaptic properties, connectivity motifs and target specificity. *Front Comput Neurosci*, 8:175.

Esposito, U. and Vasilaki, E. (Submitted 2015). Detection of multiple and overlapping bidirectional communities within large, directed and weighted networks of neurons. *Nature Scientific Reports*.

Finnerty, G. T., Roberts, L. S., and Connors, B. W. (1999). Sensory experience modifies the short-term dynamics of neocortical synapses. *Nature*, 400(6742):367–371.

Fioravante, D. and Regehr, W. G. (2011). Short-term forms of presynaptic plasticity. *Curr Opin Neurobiol*, 21(2):269–274.

Fiorillo, C. D., Tobler, P. N., and Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science*, 299(5614):1898–1902.

Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, 486(3-5):75–174.

Fortunato, S. and Barthélemy, M. (2007). Resolution limit in community detection. *Proc Natl Acad Sci U S A*, 104(1):36–41.

Frackowiak, R. and Markram, H. (2015). The future of human cerebral cartography: a novel approach. *Philos Trans R Soc Lond B Biol Sci*, 370(1668).

Friston, K. (2011). Functional and effective connectivity: A review. *Brain Connectivity*, 1(1):13–36.

Fuhrmann, G., Cowan, A., Segev, I., Tsodyks, M., and Stricker, C. (2004). Multiple mechanisms govern the dynamics of depression at neocortical synapses of young rats. *J Physiol*, 557(Pt 2):415–38.

Fuhrmann, G., Segev, I., Markram, H., and Tsodyks, M. (2002). Coding of temporal information by activity-dependent synapses. *J Neurophysiol*, 87(1):140–148.

Fusi, S., Asaad, W. F., Miller, E. K., and Wang, X.-J. (2007). A neural circuit model of flexible sensorimotor mapping: learning and forgetting on multiple timescales. *Neuron*, 54(2):319–333.

Garey, M. R. and Johnson, D. S. (1990). *Computers and Intractability: A Guide to the Theory of NP-completeness*. W. H. Freeman & Co., New York, USA.

Gerstner, W., Kempter, R., Van Hemmen, J., and Wagner, H. (1996). A neuronal learning rule for sub-millisecond temporal coding. *Nature*, 383(6595):76–78.

Gerstner, W. and Kistler, W. M. (2002a). Mathematical formulations of hebbian learning. *Biol Cybern*, 87(5-6):404–415.

Gerstner, W. and Kistler, W. M. (2002b). *Spiking Neuron Models*. Cambridge University Press.

Girvan, M. and Newman, M. E. J. (2002). Community structure in social and biological networks. *Proc Natl Acad Sci U S A*, 99(12):7821–7826.

Green, D. and Sadedin, S. (2005). Interactions matter–complexity in landscapes and ecosystems. *Ecological Complexity*, 2:117–130.

Grillner, S., Markram, H., De Schutter, E., Silberberg, G., and LeBeau, F. (2005). Microcircuits in action - from cpgs to neocortex. *Trends in Neuroscience*, 28(10):525–33.

Grossberg, S. and Versace, M. (2008). Spikes, synchrony, and attentive learning by laminar thalamocortical circuits. *Brain Res*, 1218:278–312.

Gupta, A., Wang, Y., and Markram, H. (2000). Organizing principles for a diversity of GABAergic interneurons and synapses in the neocortex. *Science*, 287(5451):273–278.

Gütig, R. and Sompolinsky, H. (2006). The tempotron: a neuron that learns spike timing-based decisions. *Nat Neurosci*, 9(3):420–428.

Guye, M., Bettus, G., Bartolomei, F., and Cozzone, P. J. (2010). Graph theoretical analysis of structural and functional connectivity mri in normal and pathological brain networks. *MAGMA*, 23(5-6):409–421.

Hai, A., Shappir, J., and Spira, M. (2010). In-cell recordings by extracellular microelectrodes. *Nat Methods*, 7:200–2.

He, B. J. (2014). Scale-free brain activity: past, present, and future. *Trends Cogn Sci*, 18(9):480–487.

He, Y. and Evans, A. (2010). Graph theoretical modeling of brain connectivity. *Curr Opin Neurol*, 23(4):341–350.

Hebb, D. O. (1949). *The Organization of Behavior: A Neuropsychological Theory*. New York: Wiley & Sons.

Hennig, M. H. (2013). Theoretical models of synaptic short term plasticity. *Front Comput Neurosci*, 7:154.

Hertz, J., Krough, A., and Palmer, R. (1991). *Introduction to the Theory of Neural Computation*. Addison Wesley, Redwood City, CA.

Hines, M., Morse, T., Migliore, M., Carnevale, N., and Shepherd, G. (2004). ModelDB: A database to support computational neuroscience. *J Comput Neurosci*, 17(1):7–11.

Hodgkin, A. and Huxley, A. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.-London*, 117(4):500–544.

Hosoi, N., Sakaba, T., and Neher, E. (2007). Quantitative analysis of calcium-dependent vesicle recruitment and its functional role at the calyx of held synapse. *J Neurosci*, 27(52):14286–14298.

Ingalhalikar, M., Parker, D., Ghanbari, Y., Smith, A., Hua, K., Mori, S., Abel, T., Davatzikos, C., and Verma, R. (2015). Connectome and maturation profiles of the developing mouse brain using diffusion tensor imaging. *Cereb Cortex*, 25(9):2696–2706.

Insel, T. R., Volkow, N. D., Li, T.-K., Battey, Jr, J. F., and Landis, S. C. (2003). Neuroscience networks: data-sharing in an information age. *PLoS Biol*, 1(1):E17.

Izhikevich, E. M. (2007). Solving the distal reward problem through linkage of stdp and dopamine signaling. *Cereb Cortex*, 17(10):2443–2452.

James, W. (1980). *The principles of psychology*. New York: Holt.

Kandell, E., Schwartz, J., and Jessel, A. (2008). *Principles of Neural Science*. McGraw-Hill.

Karrer, B., Newman, M. E. J., and Zdeborové, L. (2014). Percolation on sparse networks. *Phys. Rev. Lett.*, 113.

Klyachko, V. A. and Stevens, C. F. (2006). Excitatory and feed-forward inhibitory hippocampal synapses work synergistically as an adaptive filter of natural spike trains. *PLoS Biol*, 4(7):e207.

Kolb, B. and Gibb, R. (2011). Brain plasticity and behaviour in the developing brain. *J Can Acad Child Adolesc Psychiatry*, 20(4):265–276.

Konorski, J. (1948). *Conditioned reflexes and neuron organization*. New York: Cambridge University Press.

Koslow, S. H. and Subramaniam, S., editors (2005). *Databasing the Brain: From Data to Knowledge (Neuroinformatics)*. Wiley.

Kötter, R. (2001). Neuroscience databases: tools for exploring brain structure-function relationships. *Philos Trans R Soc Lond B Biol Sci*, 356(1412):1111–1120.

Kullmann, D. M. and Lamsa, K. P. (2007). Long-term synaptic plasticity in hippocampal interneurons. *Nat Rev Neurosci*, 8(9):687–699.

Lancichinetti, A. and Fortunato, S. (2009). Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities. *Phys Rev E Stat Nonlin Soft Matter Phys*, 80(1 Pt 2):016118.

Le Be', V. and Markram, H. (2006). Spontaneous and evoked synaptic rewiring in the neonatal neocortex. *P. Natl. Acad. Sci. USA*, 103(35):13214–9.

Lefort, S., Tomm, C., Floyd Sarria, J.-C., and Petersen, C. C. H. (2009). The excitatory neuronal network of the c2 barrel column in mouse primary somatosensory cortex. *Neuron*, 61(2):301–316.

Legenstein, R., Pecevski, D., and Maass, W. (2008). A learning theory for reward-modulated spike-timing-dependent plasticity with application to biofeedback. *PLoS Comput Biol*, 4(10):e1000180.

Levina, A., Herrmann, J. M., and Geisel, T. (2009). Phase transitions towards criticality in a neural system with adaptive interactions. *Phys Rev Lett*, 102(11):118110.

Lichtman, J., Livet, J., and Sanes, J. (2008). A technicolour approach to the connectome. *Nat Rev Neurosci*, 9:417–22.

Lichtman, J. W. and Sanes, J. R. (2008). Ome sweet ome: what can the genome tell us about the connectome? *Curr Opin Neurobiol*, 18(3):346–353.

Lømo, T. (1964). Frequency potentiation of excitatory synaptic acttivity in the dendate area of the hippocampal formation. *Acta Physiol. Scand.*, 68:128.

Lømo, T. (2003). The discovery of long-term potentiation. *Philos Trans R Soc Lond B Biol Sci*, 358(1432):617–620.

López-Muñoz, F., Boya, J., and Alamo, C. (2006). Neuron theory, the cornerstone of neuroscience, on the centenary of the nobel prize award to santiago ramón y cajal. *Brain Res Bull*, 70(4-6):391–405.

Lugaro, E. (1913). *Modern Problems in Psychiatry*. The University Press, Manchester.

Luo, L., Callaway, E. M., and Svoboda, K. (2008). Genetic dissection of neural circuits. *Neuron*, 57(5):634–660.

Lynch, G. S., Dunwiddie, T., and Gribkoff, V. (1977). Heterosynaptic depression: a postsynaptic correlate of long-term potentiation. *Nature*, 266(5604):737–739.

Maass, W. and Markram, H. (2002). Synapses as dynamic memory buffers. *Neural Netw*, 15(2):155–161.

Malliaros, F. D. and Vazirgiannis, M. (2013). Clustering and community detection in directed networks: A survey. *Physics Reports*, 533(4).

Mariani, J., Simonini, M. V., Palejev, D., Tomasini, L., Coppola, G., Szekely, A. M., Horvath, T. L., and Vaccarino, F. M. (2012). Modeling human cortical development in vitro using induced pluripotent stem cells. *Proc Natl Acad Sci U S A*, 109(31):12770–12775.

Markram, H., Gerstner, W., and Sjöström, P. J. (2011). A history of spike-timing-dependent plasticity. *Front Synaptic Neurosci*, 3:4.

Markram, H., Lübke, J., Frotscher, M., and Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic aps and epsps. *Science*, 275(5297):213–215.

Markram, H., Pikus, D., Gupta, A., and Tsodyks, M. (1998a). Potential for multiple mechanisms, phenomena and algorithms for synaptic plasticity at single synapses. *Neuropharmacology*, 37(4-5):489–500.

Markram, H. and Tsodyks, M. (1996). Redistribution of synaptic efficacy between neocortical pyramidal neurones. *Nature*, 382:807–9.

Markram, H., Wang, Y., and Tsodyks, M. (1998b). Differential signaling via the same axon of neocortical pyramidal neurons. *Proc. Nat. Acad. Sci. U.S.A.*, 95(5323-8).

Matveev, V. and Wang, X. J. (2000). Differential short-term synaptic plasticity and transmission of complex spike trains: to depress or to facilitate? *Cereb Cortex*, 10(11):1143–1153.

Mechelli, A., Price, C. J., Friston, K. J., and Ishai, A. (2004). Where bottom-up meets top-down: neuronal interactions during perception and imagery. *Cereb Cortex*, 14(11):1256–1265.

Mills, K. L. and Tamnes, C. K. (2014). Methods and considerations for longitudinal structural brain imaging analysis across development. *Dev Cogn Neurosci*, 9:172–190.

Minderer, M., Liu, W., Sumanovski, L., Kügler, S., Helmchen, F., and Margolis, D. (2012). Chronic imaging of cortical sensory map dynamics using a genetically encoded calcium indicator. *J Physiol*, 590:99–107.

Morrison, A., Diesmann, M., and Gerstner, W. (2008). Phenomenological models of synaptic plasticity based on spike timing. *Biological Cybernetics*, 98(6):459–78.

Muzio, L. and Consalez, G. G. (2013). Modeling human brain development with cerebral organoids. *Stem Cell Res Ther*, 4(6):154.

Natschläger, T., Maass, W., and Zador, A. (2001). Efficient temporal processing with biologically realistic dynamic synapses. *Network*, 12(1):75–87.

Newman, M. (2004). Detecting community structure in networks. *The European Physical Journal B*, 38(2):321–330.

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

Newman, M. E. J. (2006). Modularity and community structure in networks. *Proc Natl Acad Sci U S A*, 103(23):8577–8582.

Newman, M. E. J. (2013). Spectral methods for network community detection and graph partitioning. *Phys. Rev. E*.

Newman, M. E. J. and Girvan, M. (2004). Finding and evaluating community structure in networks. *Phys Rev E Stat Nonlin Soft Matter Phys*, 69(2 Pt 2):026113.

Ovelgönne, M. and Geyer-schulz, A. (2012). An ensemble learning strategy for graph clustering. In *10th DIMACS Implementation Challenge Graph Partitioning and Graph Clustering*.

Palla, G., Derényi, I., Farkas, I., and Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043):814–818.

Papadimitriou, C. H. (1977). The euclidean travelling salesman problem is np-complete. *Theoretical Computer Science*, 4(3):237–244.

Perin, R., Berger, T., and Markram, H. (2011). A synaptic organizing principle for cortical neuronal groups. *Proc. Nat. Acad. Sci. U.S.A.*, 108(13):5419–24.

Pernice, V. and Rotter, S. (2013). Reconstruction of sparse connectivity in neural networks from spike train covariances. *Journal of Statistical Mechanics: Theory and Experiment*.

Pfister, J.-P., Dayan, P., and Lengyel, M. (2010). Synapses with short-term plasticity are optimal estimators of presynaptic membrane potentials. *Nat Neurosci*, 13(10):1271–1275.

Pfister, J.-P. and Gerstner, W. (2006). Triplets of spikes in a model of spike timing–dependent plasticity. *J Neurosci*, 26:9673–82.

Pignatelli, M. (2009). *Structure and Function of the Olfactory Bulb Microcircuit*. PhD thesis, École Polytechnique Fédérale de Lausanne, http://library.epfl.ch/en/theses/?nr=4275.

Porter, M., Onnela, J.-P., and Mucha, P. J. (2009). Communities in networks. *Notices of the American Mathematical Society*, 56(9):1082–1097.

Ramón y Cajal, S. (1954). *Neuron theory or reticular theory? Objective evidence of the anatomical unity of nerve cells*. Madrid, Consejo Superior de Investigaciones Científicas, Instituto Ramón y Cajal.

Regehr, W. G. (2012). Short-term presynaptic plasticity. *Cold Spring Harb Perspect Biol*, 4(7):a005702.

Reyes, A., Lujan, R., Rozov, A., Burnashev, N., Somogyi, P., and Sakmann, B. (1998). Target-cell-specific facilitation and depression in neocortical circuits. *Nat Neurosci*, 1(4):279–285.

Reyes, A. and Sakmann, B. (1999). Developmental switch in the short-term modification of unitary epsps evoked in layer 2/3 and layer 5 pyramidal neurons of rat neocortex. *J Neurosci*, 19(10):3827–35.

Richmond, P., Buesing, L., Giugliano, M., and Vasilaki, E. (2011). Democratic population decisions result in robust policy-gradient learning: A parametric study with GPU simulations. *PLoS One*, 6(5):e18539.

Rinaldi, T., Perrodin, C., and Markram, H. (2008). Hyper-connectivity and hyper-plasticity in the medial prefrontal cortex in the valproic acid animal model of autism. *Front Neural Circuits*, 2:4.

Romani, A., Marchetti, C., Bianchi, D., Leinekugel, X., Poirazi, P., Migliore, M., and Marie, H. (2013). Computational modeling of the effects of amyloid-beta on release probability at hippocampal synapses. *Front Comput Neurosci*, 7:1.

Rosvall, M. and Bergstrom, C. T. (2007). An information-theoretic framework for resolving community structure in complex networks. *Proc Natl Acad Sci U S A*, 104(18):7327–7331.

Rotman, Z. and Klyachko, V. A. (2013). Role of synaptic dynamics and heterogeneity in neuronal learning of temporal code. *J Neurophysiol*, 110(10):2275–2286.

Rumelhart, D. E., McClelland, J. L., and Group, P. R., editors (1986). *Parallel Distributed Processing*. MIT Press.

Schultz, W., Dayan, P., and Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, 275(5306):1593–1599.

Schummers, J., Sharma, J., and Sur, M. (2005). Bottom-up and top-down dynamics in visual cortex. *Prog Brain Res*, 149:65–81.

Sejnowski, T. J. and Rosenberg, C. R. (1987). Parallel networks that learn to pronounce english text. *COMPLEX SYSTEMS*.

Seung, H. (2009). Reading the book of memory: Sparse sampling reading the book of memory: Sparse sampling versus dense mapping of connectomes. *Neuron*, 62:17–29.

Sherrington, C. (1897). *A textbook of physiology. Part III: The central nervous system*. Macmillan and Co, London.

Silberberg, G., Grillner, S., LeBeau, F., Maex, R., and Markram, H. (2005). Synaptic pathways in neural microcircuits. *Trends in Neuroscience*, 28(10):541–51.

Silberberg, G. and Markram, H. (2007). Disynaptic inhibition between neocortical pyramidal cells mediated by martinotti cells. *Neuron*, 53(5):735–46.

Sjöström, P., Turrigiano, G., and Nelson, S. (2001). Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. *Neuron*, 32:1149–1164.

Sjöström, P. J., Turrigiano, G. G., and Nelson, S. B. (2003). Neocortical ltd via coincident activation of presynaptic nmda and cannabinoid receptors. *Neuron*, 39(4):641–654.

Sjöström, P. J., Turrigiano, G. G., and Nelson, S. B. (2007). Multiple forms of long-term plasticity at unitary neocortical layer 5 synapses. *Neuropharmacology*, 52(1):176–184.

Smith, S. J. (2007). Circuit reconstruction tools today. *Curr Opin Neurobiol*, 17(5):601–608.

Song, S. and Abbott, L. (2001). Cortical development and remapping through spike timing-dependent plasticity. *Neuron*, 32:339–50.

Song, S., Miller, K., and Abbott, L. (2000). Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nat Neurosci*, 3:919–26.

Song, S., Sjöström, P., Reigl, M., Nelson, S., and Chklovskii, D. (2005). Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biol*, 3(3):e68.

Sporns, O. (2011a). The human connectome: a complex network. *Ann N Y Acad Sci*, 1224:109–125.

Sporns, O. (2011b). The non-random brain: efficiency, economy, and complex dynamics. *Front Comput Neurosci*, 5:5.

Sporns, O. (2013). Structure and function of complex brain networks. *Dialogues Clin Neurosci*, 15(3):247–262.

Sporns, O., Chialvo, D. R., Kaiser, M., and Hilgetag, C. C. (2004). Organization, development and function of complex brain networks. *Trends Cogn Sci*, 8(9):418–425.

Sporns, O., Tononi, G., and Kötter, R. (2005). The human connectome: A structural description of the human brain. *PLoS Comput Biol*, 1(4):e42.

Stent, G. S. (1973). A physiological mechanism for hebb's postulate of learning. *Proc Natl Acad Sci U S A*, 70(4):997–1001.

Stevens, C. F. and Wesseling, J. F. (1998). Activity-dependent modulation of the rate at which synaptic vesicles become available to undergo exocytosis. *Neuron*, 21(2):415–424.

Szatmáry, B., Richer, M., Nageswaran, J. M., Petre, C., Piekniewski, F., Sokol, S., , and Izhikevich, E. M. (2013). Emergence of bottom-up saliency in a spiking model of v1. *BMC Neuroscience*, 14.

Tansey, E. M. (1997). Not committing barbarisms: Sherrington and the synapse, 1897. *Brain Res Bull*, 44(3):211–212.

Tanzi, E. (1893). I fatti e le induzioni dell' odierna istologia del sistema nervoso. *Riv Sper Fren Med Leg*, 19:419–472.

Testa-Silva, G., Loebel, A., Giugliano, M., de Kock, C. P. J., Mansvelder, H. D., and Meredith, R. M. (2012). Hyperconnectivity and slow synapses during early development of medial prefrontal cortex in a mouse model for mental retardation and autism. *Cereb Cortex*, 22(6):1333–1342.

Tetzlaff, C., Kolodziejski, C., Markelic, I., and Wörgötter, F. (2012). Time scales of memory, learning, and plasticity. *Biol Cybern*, 106(11-12):715–726.

Thomson, A. M. (2000). Facilitation, augmentation and potentiation at central synapses. *Trends Neurosci*, 23(7):305–312.

Thorndike, E. L. (1911). *Animal Intelligence: Experimental Studies*. Macmillan.

Tobler, P. N., Fiorillo, C. D., and Schultz, W. (2005). Adaptive coding of reward value by dopamine neurons. *Science*, 307(5715):1642–1645.

Toga, A. W., Clark, K. A., Thompson, P. M., Shattuck, D. W., and Van Horn, J. D. (2012). Mapping the human connectome. *Neurosurgery*, 71(1):1–5.

Tsodyks, M. and Markram, H. (1997). The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proc. Nat. Acad. Sci. U.S.A.*, 94:719–23.

Tsodyks, M., Pawelzik, K., and Markram, H. (1998). Neural networks with dynamic synapses. *Neural Comp*, 10:821–35.

Urbanczik, R. and Senn, W. (2009). Reinforcement learning in populations of spiking neurons. *Nat Neurosci*, 12(3):250–252.

Van Bussel, F., Kriener, B., and Timme, M. (2011). Inferring synaptic connectivity from spatio-temporal spike patterns. *Front Comput Neurosci*, 5:3.

Van Essen, D. C. and Ugurbil, K. (2012). The future of the human connectome. *Neuroimage*, 62(2):1299–1310.

Van Essen, D. C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T. E. J., Bucholz, R., Chang, A., Chen, L., Corbetta, M., Curtiss, S. W., Della Penna, S., Feinberg, D., Glasser, M. F., Harel, N., Heath, A. C., Larson-Prior, L., Marcus, D., Michalareas, G., Moeller, S., Oostenveld, R., Petersen, S. E., Prior, F., Schlaggar, B. L., Smith, S. M., Snyder, A. Z., Xu, J., Yacoub, E., and , W. U.-M. H. C. P. C. (2012). The human connectome project: a data acquisition perspective. *Neuroimage*, 62(4):2222–2231.

van Rossum, M. C., Bi, G. Q., and Turrigiano, G. G. (2000). Stable hebbian learning from spike timing-dependent plasticity. *J Neurosci*, 20(23):8812–8821.

Varela, J., Sen, K., Gibson, J., Fost, J., Abbott, L., and Nelson, S. (1997). A quantitative description of short-term plasticity at excitatory synapses in layer 2/3 of rat primary visual cortex. *J Neurosci*, 17(20):7926–40.

Varshney, L. R., Chen, B. L., Paniagua, E., Hall, D. H., and Chklovskii, D. B. (2011). Structural properties of the caenorhabditis elegans neuronal network. *PLoS Comput Biol*, 7(2):e1001066.

Vasilaki, E., Frémaux, N., Urbanczik, R., Senn, W., and Gerstner, W. (2009a). Spike-based reinforcement learning in continuous state and action space: when policy gradient methods fail. *PLoS Comput Biol*, 5(12):e1000586.

Vasilaki, E., Fusi, S., Wang, X.-J., and Senn, W. (2009b). Learning flexible sensori-motor mappings in a complex network. *Biol Cybern*, 100(2):147–158.

Vasilaki, E. and Giugliano, M. (2012). Emergence of connectivity patterns from long-term and short-term plasticities. In Villa, A., Duch, W., Érdi, P., Masulli, F., and Palm, G., editors, *Artificial Neural Networks and Machine Learning – ICANN 2012*, volume 7552 of *Lecture Notes in Computer Science*, pages 193–200. Springer Berlin Heidelberg.

Vasilaki, E. and Giugliano, M. (2014). Emergence of connectivity motifs in networks of model neurons withshort- and long-term plastic synapses. *PLoS One*, 9(1):e84626.

Vogels, T. P. and Abbott, L. F. (2009). Gating multiple signals through detailed balance of excitation and inhibition in spiking networks. *Nat Neurosci*, 12(4):483–491.

Vogels, T. P., Sprekeler, H., Zenke, F., Clopath, C., and Gerstner, W. (2011). Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks. *Science*, 334(6062):1569–1573.

Wang, J., Zuo, X., Dai, Z., Xia, M., Zhao, Z., Zhao, X., Jia, J., Han, Y., and He, Y. (2013). Disrupted functional brain connectome in individuals at risk for alzheimer's disease. *Biol Psychiatry*, 73(5):472–481.

Wang, L. Y. and Kaczmarek, L. K. (1998). High-frequency firing helps replenish the readily releasable pool of synaptic vesicles. *Nature*, 394(6691):384–388.

Wang, Y., Markram, H., Goodman, P., Berger, T., Ma, J., and Goldman-Rakic, P. (2006). Heterogeneity in the pyramidal network of the medial prefrontal cortex. *Nat Neurosci*, 9(4):534–42.

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

Wedeen, V., Rosene, D., Wang, R., Dai, G., Mortazavi, F., Hagmann, P., Kaas, J., and Tseng, W. (2012). The geometric structure of the brain fiber pathways. *Science*, 335(6076):1628–34.

Wegener, I. (2005). *Complexity Theory: Exploring the Limits of Efficient Algorithms*. Springer Science & Business Media.

White, J. G., Southgate, E., Thomson, J. N., and Brenner, S. (1986). The structure of the nervous system of the nematode caenorhabditis elegans. *Philos Trans R Soc Lond B Biol Sci*, 314(1165):1–340.

Wickersham, I., Lyon, DC Barnard, E., Mori, T., Finke, S., Conzelmann, K., Young, J., and Callaway, E. (2007). Monosynaptic restriction of transsynaptic tracing from single, genetically targeted neurons. *Neuron*, 53(5):639–47.

Yu, S., Huang, D., Singer, W., and Nikolic, D. (2008). A small world of neuronal synchrony. *Cereb Cortex*, 18(12):2891–2901.

Zhang, F., Aravanis, A., Adamantidis, A., de Lecea, L., and Deisseroth, K. (2007). Circuit-breakers: optical technologies for probing neural signals and systems. *Nat Rev Neurosci*, 8:577–81.

Zhou, J., Gennatas, E. D., Kramer, J. H., Miller, B. L., and Seeley, W. W. (2012). Predicting regional neurodegeneration from the healthy brain functional connectome. *Neuron*, 73(6):1216–1227.

Zucker, R. and Regehr, W. (2002). Short-term synaptic plasticity. *Annu Rev Physiol.*, 64:355–405.