UNIVERSITETET
I OSLO

Master thesis

Analysis of the Functional Role of Directed Simplicial Structures in Biological Neural Networks

Sara Pernille Jensen

Master in Computational Science: Physics
60 credits

Department of Physics
Faculty of Mathematics and Natural Sciences
Spring 2023

In memory of my family

Abstract

In this project we analysed the functional role of directed simplicial structures in biological neural networks. This was inspired by previous work analysing the correlations in the activity in such networks, but a novel approach was here taken, emphasising the causal structure and dependencies in the dynamics of the networks. The motivation for such an approach was the belief that such a causal description more strongly reflect the underlying causal mechanisms of the systems, such that detecting these are more conducive to a complete understanding of the system at multiple levels. The datasets considered were computationally generated neural activity in a variety of networks with different simplicial structures. It was then tested whether the low-level functions considered were dependent on the size and completeness of the directed simplices. Additionally, we investigated whether the simplicies formed coherent functional groups with regard to the metrics chosen. As well as testing for specific functions, a model-independent approach was also tested, where we attempted to train machine learning models to predict the simplicial structures of the networks from their activity.

From the results obtained, we were unable to attribute any functional properties to the simplicial structures, so the question of whether these structures serve any clear functional role remains unanswered. Further, we found a very high within-group variance with respect to the different functions considered, indicating that grouping the simplices together based on their size and completeness is not informative of the functional properties of the activity in these networks. This led to a more critical discussion regarding the methodological approach taken in this project, as well as the pursuitworthiness of such simplicial structures in future neuroscientific research.

Contents

1 Introduction

With its approximately 86 billion neurons, the human brain remains one of the least understood systems of modern science. Much progress has been made in the reductionist program, such that we largely understand the dynamics of individual neurons and their interactions. However, we are still very far from any understanding of how the network structures and activity of these neurons give rise to the complex behaviour of humans and other animals. In recent decades, an increasingly popular and fruitful approach in neuroscience has been to study the brain in terms of the organisation and function of brain networks. Analysing the networks of the brain at different scales and using different measures, this has led to new methods for diagnosis of neural diseases [55], better understanding of cognitive features [37] as well as new conceptual frameworks for understanding both low-level dynamics and cognition [53]. This subdiscipline of neuroscience is what has become known as *connectomics*, under which category this project falls.

The networks of the brain can be analysed at different levels, but in the current project, the emphasis has been on networks at the level of individual neurons and synapses. The overarching motivation has here been to explain the functionality of these structures. A necessary question is thus what such an explanation would involve.

In the past, many of our theories of scientific explanations have been based on physics as the "model science", and have consequently assumed a rather reductionist view of nature. This assumption might well hold in physics, but does not necessarily generalise to other disciplines, such as biology. In response to this, neuroscientist David Marr presented a more suitable framework for describing what constitutes understanding of complex and hierarchical biological systems. He argued that a full understanding of biological systems requires understanding at three different levels: the computational, the algorithmic and the implementational [36]. The computational level concerns the purpose of the computation, i.e. the problem it is trying to solve. The algorithmic level describes the algorithm used to solve the problem. Here, the states of the system are seen as representational of different informational states, and the algorithm describes the information processing in terms of input and output. Finally, the implementational level describes how these algorithms are instantiated physically in the system. Each level is thus the realisation of the preceding one. Further, in complex and hierarchical systems, such as the brain, these three levels can in turn exist at different levels of description, such that what constitutes the holistic computation at one level might also be the smallest element of the implementation at the level above. Marr argued that a problem in much of neuroscience was that many scientists operated at only one of the levels, without regard for the other two. This gives an incomplete description, insufficient for providing a holistic understanding of the system, which requires all three.

A much discussed problem in neuroscience today is that we are in a situation where observational "big-data" is abundant, yet the organising principles required to interpret this data in meaningful ways and thereby turning it into understanding are still lacking. A common approach in the field is to look for correlations in the data at different levels of description. However, as will be argued in later sections, there is much to be said against the explanatory potential of mere descriptive accounts of correlations, since these do not necessarily reflect any underlying causal structure of the systems. Most importantly, whereas causal relations are believed to reflect the objective ontological structure of the system,

descriptions of correlations are epistemic in that they only reflect patterns in the emergent activity, rather than the underlying mechanisms constraining said activity. This is one motivation for a turn towards causal explanations in neuroscience, as the current project is an example of.

Given this focus on causality, it is reasonable that the structure of choice should be one which has the potential to be treated as causal units. The set of network structures investigated here are so-called *directed simplices* of different sizes. These are all-to-all connected cliques of neurons with a single source and a single sink, such that the flow of activity through the structure is uni-directional. This is a network structure which has only very recently been shown any interest in neuroscience, such as in [47] and [21], where they hypothesise that simplicial structures in the functional and structural connectivity of the brain might be related to certain higher-level cognitive functions. As per usual, both of these studies focused on correlations in the activity of the networks, and to the best of our knowledge, this project constitutes the first attempt to study the emergent functionality of simplicial structures using the framework of causal inference.

The datasets analysed were generated computationally, using a simulator of neural activity based on a generalised linear model, and the aim was to study the dependency of the resulting activity on the simplicial network structures. The main benefits of using simulated data is that all information about the system is available at all times, and it is possible to include interventions, thereby facilitating the identification of causal dependencies between the components of the networks. In terms of Marr's explanatory levels, one can consider the set of network structures and the neuronal dynamics between them as the implementational level, which was given in our case. The aim of this project was to connect these descriptions with their corresponding algorithms, which would describe the flow of information in the systems. In other words, the goal was to find the functional role of a certain set of network structures at the circuit level. The overarching goal would of course be to further connect this level of analysis with the computational description, but this would require an understanding of the *purpose* of the algorithms, which would in turn require knowledge of how the higher-level cognitive functions supervene on the low-level circuit functions here studied. Unfortunately, these bridge principles are as yet unknown. Thus, we were limited to looking at the neural dynamics at the level of neurons and their activity, and the best we could hope for was to find some lower-level functional properties which might plausibly be connected with higher-level cognitive functions, and which might thereby serve as proxies for these. The choice of low-level features and metrics will be discussed in detail in later sections.

Finally, inspired by recent successes in applying machine learning models to datasets to discover unknown functional mappings, most notably AlphaFold [16], a similar approach was attempted here. It was tested whether a graph neural network, which had successfully been employed for causal inference on the same datasets, was able to discern some functional features of the network activity and thereby reconstruct the simplicial structure from this. However, the main emphasis was on the analysis of the simplices' functional role, for, to quote Krishna Shenoy: "[w]hile it is tempting to [...] treat this unique neural dataset as just another "big data" dataset and unleash somewhat generic machine learning algorithms on it, [this] would likely limit the full extent of insights that are believed to be possible" [59, p. 88]. More specifically, a positive result with such a method would simply be an indication that some function exists, though we would be none the wiser as to what it is.

Thus, since the primary aim of this project was to gain further understanding rather than just accurate predictions, the main emphasis has been on trying to establish the functional role of simplicial structures in biological neural networks.

1.1 Research Questions

The main research questions in this project were as follows

- Do directed simplices in biological neural networks of different sizes form meaningful functional groups as well as structural groups?
- Are larger directed simplices more functionally significant than smaller ones?
- Can directed simplices meaningfully be treated as causal units in the networks, and do their causal significance increase with simplex size?

1.2 Outline of Thesis

In section 2, an overview of the theoretical background from the different related fields is presented. The datasets and the main statistical methods used are also presented here. In section 3, we present a more detailed account of the research purpose and the motivation for the approach adopted in this project. This involves some more philosophical perspectives on the methodology and the conceptual challenges related to this, along with more details of the different metrics chosen to answer the research questions. Next, the details and results from the functional analysis of the simplicial structures is presented in section 4. This section is in turn divided into three main parts, one for each of the metrics considered. The different metrics are presented at the beginning of each subsection, and each ends with a brief discussion of the results, in addition to a short summary and discussion at the end of the main section. The next section, section 5, is where the machine learning model is presented, along with some motivating remarks regarding its use. The results of this investigation is presented at the end of the section, along with a discussion of these. Finally, a summary and conclusion of all the findings is presented in section 6, along with a discussion regarding the limitations of the research. This section also includes some reflections on how similar questions might be pursued in future research.

2 Background and Theory

This rather interdisciplinary project brings together research from a range of different fields. As a result, a somewhat extensive background section is required.

The first section covers the neuroscientific background, both the very basics necessary to understand the dynamical systems studied here, as well as a more extensive section on connectomics. Though not strictly linked with the current project, this part is included for a fuller understanding of the motivation for the project, as it highlights many of the practical and experimental challenges currently being grappled with in the field. The same holds for the section on causal inference.

The other sections are more directly linked with the current research, and we have aimed to present these more laconically. First, we give the mathematical background regarding the simplicial structures. Second, an introduction is given to the field of causal inference and its significance. Third, a brief overview of the theory behind the machine learning methods used in the project is given. Fourth, we present the different datasets investigated, along with a more detailed explanation of the simulator used to generate these. Finally, an outline of the different statistical tools used for the data analysis is given.

2.1 Neuroscience

The overarching aim of this project has been to investigate and analyse the causal structure of biological neural networks. This naturally builds on knowledge from many fields in neuroscience, so it will be useful to have decent overview of the most relevant topics. First, a brief description of the basic constituents of biological neural networks and their dynamics is provided. Second, given that the current project falls under the category of connectomics, an outline of the different branches of connectomics and their related challenges will be presented. A fuller exposition of the model used to generate the artificial data used in this project can be found in section 2.5.2.

2.1.1 Biological Neural Networks - The Basics

The Neuron Neurons are usually considered the fundamental processing units of the brain. They come in different shapes and sizes, but they all consist of three functionally distinct parts: the soma, the dendrites, and the axon. The soma is the cell body where the information is processed and the action signal originates. The dendrites are the branching structures near the soma. These are connected to other neurons through synapses, and are where the signals from these other neurons enter before they are transmitted to the soma. The soma processes the incoming signals, and if the total incoming activity exceeds some approximate threshold, a signal is generated in the axon hillock in the soma and passed along the *axon*. This in turn has many outgoing branches, known as *axon terminals*, connecting it to the dendrites of other neurons, which then receive the signal. A figure of two connected neurons with their functional parts is shown in fig. 1.

Both the intracellular (the cytoplasm) and the extracellular media of neurons contains a range of different ions. Normally, the concentration of negatively charged ions are higher on the inside than the outside, giving a membrane potential of about -70 mV in mammals. This

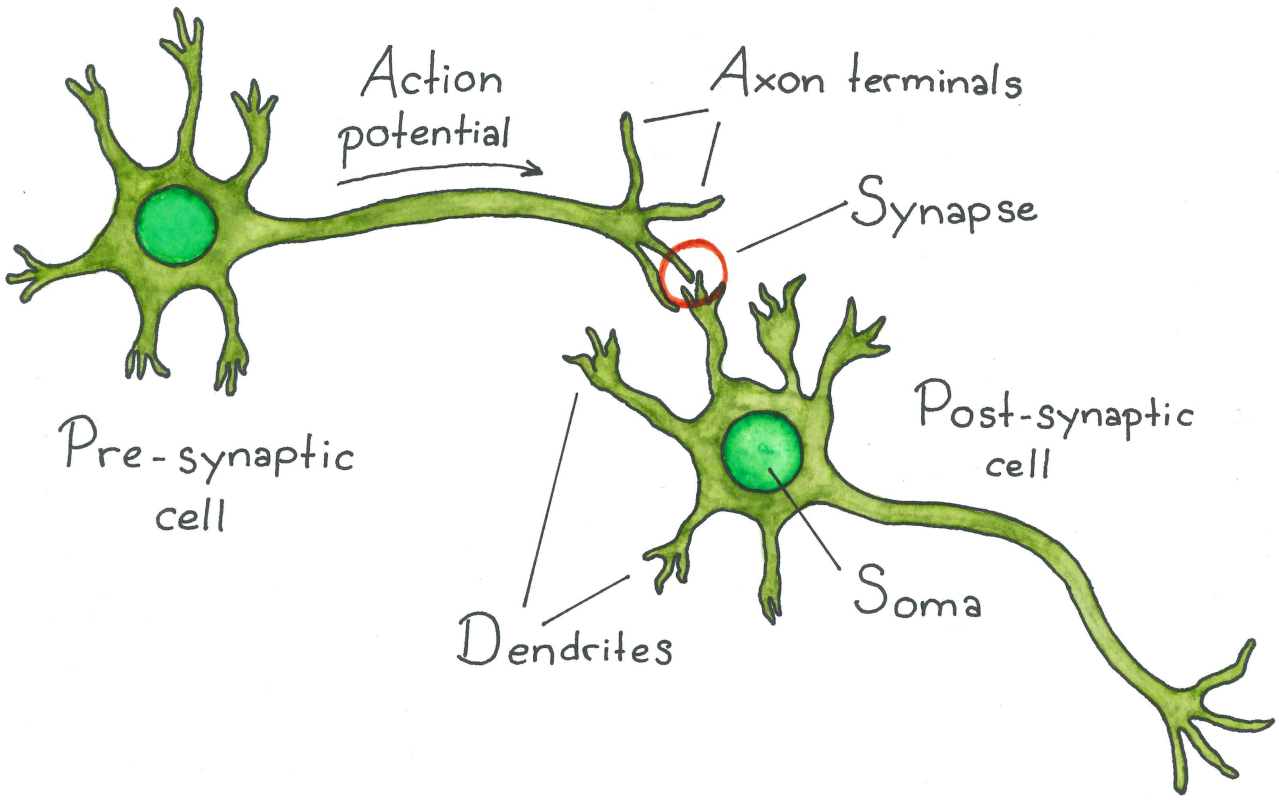


Figure 1: Figure of two neurons, here labelled as cells, connected together via an axon and a synapse. The action potential goes from the pre-synaptic to the post-synaptic neuron, and is transmitted at the synapse.

is known as the *resting membrane potential*. Changes in the concentration of different ions in the cytoplasm leads to changes in the membrane potential, and this is what ultimately triggers an action potential to be fired. Such changes are regulated by a range of different ion channels in the membrane, through which different types of ions may enter or exit the neuronal membrane.

The Action Potential The signals transmitted by neurons come in the form of a series of electrical pulses, each of which is called an *action potential* or a *spike*. Action potentials last for about 1-2 ms, and they all have the same amplitude. Since all action potentials are fairly equal in shape and strength, they do not in themselves encode any information. Rather, the information is encoded in the *spike trains*, i.e. the series of subsequent spikes, including their timing and the number of spikes. An illustration of the shape and the different phases of the firing of an action potential is shown in fig. 2.

Whether a spike is generated in the soma is dependent on the state and membrane potential of the neuron. The membrane potential is affected by incoming signals which are passed on through the dendrites. These can either make the membrane potential more negative (*hyperpolarised*) or less negative (*depolarised*). An action potential is triggered by an increase in the membrane potential (depolarisation), up to an approximate threshold. The

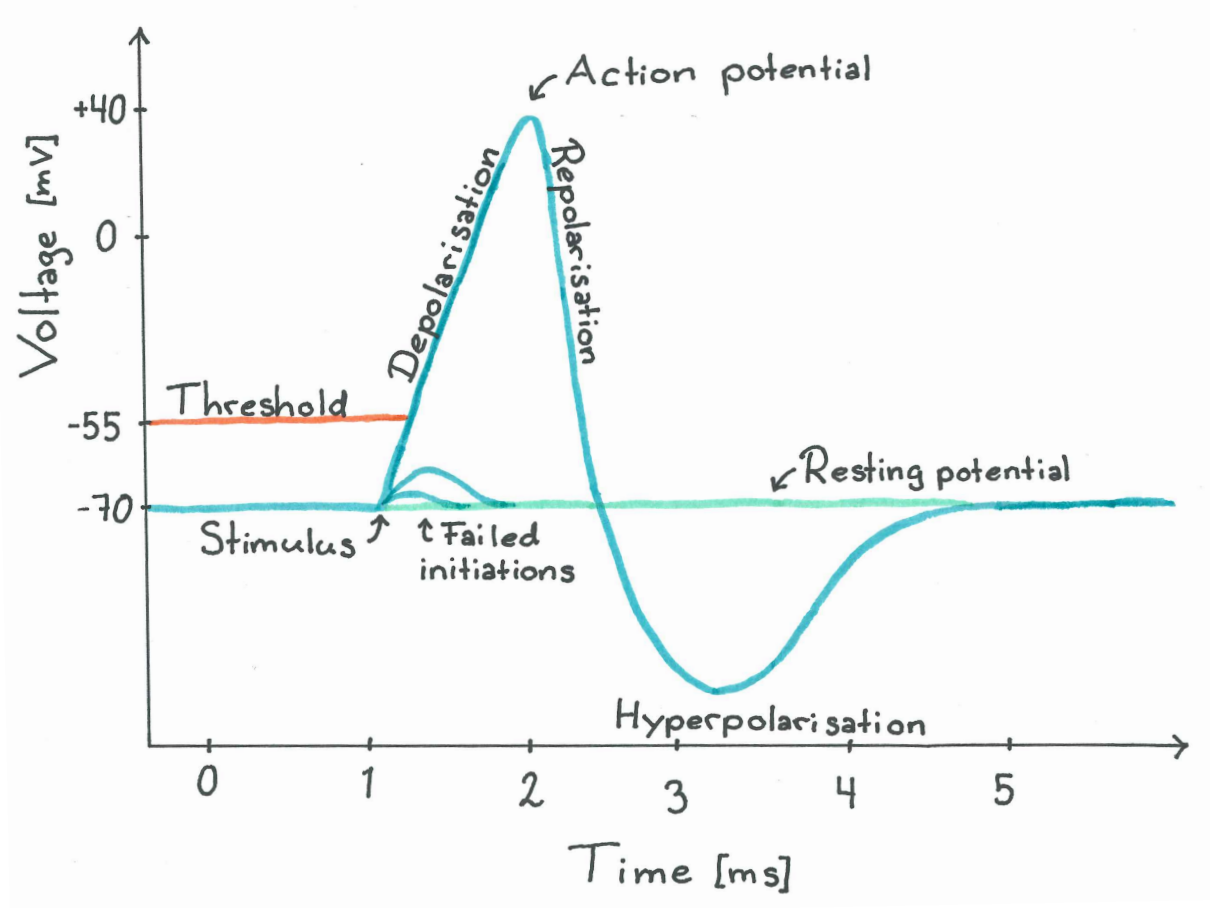


Figure 2: Diagram of a single action potential, with membrane potential plotted as function of time. After a stimulus with sufficient strength to increase the voltage above the threshold, the potential is first depolarised, before it reaches the peak, and becomes repolarised, until it reaches a minimum, labelled hyperpolarisation. After this, the membrane potential returns to its stable state of the resting potential. Each change is triggered by changes in the intracellular and extracellular ion concentration, though these are not included here.

action potential itself is characterised by a sharp increase in the membrane potential, usually to a positive value, followed by a less sharp decrease in the membrane potential, known as *repolarisation*, to a membrane potential below the resting potential. Immediately after an action potential has been fired, it is impossible for a new one to be generated, regardless of the incoming signals. The minimal period between two action potentials for a neuron is known as the *absolute refractory period*. This ensures that the spikes are usually distinct and well separated. After the absolute refractory period, there is a *relative refractory period*, during which the soma requires far stronger input to elicit a new action potential than what is normally the case [20].

The Synapse When an action potential is transmitted from one neuron to another, it is said to be sent from the *pre-synaptic* to the *post-synaptic* neuron. The synapse is thus the point at which the two neurons meet and the action potential is transmitted. There

are two different types of synapses, chemical and electrical, where the chemical synapse is the most common type in vertebrate brains, constituting about 90 % of the synapses[4], and is also better understood [62]. At a chemical synapse, there is a small gap (a *synaptic cleft*) between the membranes of the pre- and post-synaptic cells. When the action potential reaches the synapse, it triggers a response leading to the release of neurotransmitters from the pre-synaptic terminal into the synaptic cleft. These are in turn taken up by receptors at the post-synaptic site, leading to ion channels opening and ions flowing into the recipient neuron. These ions may be positive or negative, depending on the neurotransmitters released. The strength of a synapse is defined as the average voltage produced in the post-synaptic neuron as a result of the action potential at the pre-synaptic neuron [41]. This is determined both by the amount of neurotransmitters released at the pre-synaptic terminal and by the sensitivity of the receptors at the post-synaptic site. The strength of a synapse may change over time, both at long and short time-scales. Neurons can be either excitatory or inhibitory, meaning that they either increase the membrane potential of their recipients, making an action potential more likely, or decrease it, making it less likely that the recipient neuron will fire. Dale's law tells us that neurons are always either excitatory or inhibitory, meaning that the neuron has the same effect on all of its post-synaptic neighbours. About 80% of the neurons in the brain are excitatory, and the remaining 20% inhibitory [62]. Importantly, chemical synapses are uni-directional, so the information always flows from the pre-synaptic to the post-synaptic neuron. Electrical synapses, on the other hand, are bi-directional. These are also known as *gap junctions*, and allow for direct transmission of electrical signals between neurons, without the use of neurotransmitters. In this project only chemical synapses will be considered, as these are the focus of most of the research in neuroscience.

Network Motifs As well as individual neurons, it is possible to treat larger circuits as functional units of the brain. Such circuits are known as *motifs*, and are defined by the nodes and edges between them, giving sets of sub-graphs with specific topological configurations. If a certain motif is found more frequently in neural networks than what one would expect from chance, the assumption is that the motif performs some significant functional role, and can therefore be considered an important processing unit of the network [19]. Formally, a network motif is only defined by its number of constituent neurons and the number of edges between these. For each motif, there will be a number of possible configurations of the edges, and the set of all such possible configurations of a given motif is known as the *motif class*. When performing the analysis, it is customary to look at the frequency of each individual class of the same motif separately, as these might have very different structural properties. Such motif analysis is a sub-discipline of the broader field of *connectomics*.

2.1.2 Connectomics

Many of the underlying premises and motivations for the current project comes from research in connectomics, so a general overview of this field and its main research questions and methodology will be of use. Connectomics is a grouping of a range of different network-centred approaches to studying the brain which is becoming increasingly prevalent in neuroscience. It is most often associated with the study of the structural connectivity map of the brain or smaller parts thereof, but more generally, one can distinguish between three

main branches of connectomics, namely structural, functional and effective. The first is the study of the static structural connectivity. This involves determining the direct anatomical connections between neurons or regions of the brain. Secondly, one can study the functional connectivity between regions or individual neurons in the brain, though regions are most commonly the unit of study here. This connectivity is a description of the statistical correlations or covariances of activity in the units under consideration, so is undirected. Finally, effective connectivity aims to go beyond mere correlations and to establish causal connections between the units, resulting in a directed and weighted connectivity diagram. In addition to different types of connectomics, the connectivity can be studied at different scales. The four primary ones are the macroscale, which looks at the connections between regions of the brain, the mesoscale, looking at the connections between different types of neurons, the microscale, which looks at the synaptic connections between individual neurons, and the nanoscale, which looks at the connections at synapses. In this project only the microscale was considered.

There is a significant amount of literature dedicated to discussing which of these methods are most conducive to understanding the functioning of the brain. Here we shall give an overview of the main research questions, methodology and limitations of each of the methods.

Structural Connectomics When the topic of connectomics is discussed, what is often meant is the study of the anatomical connections of the brain, commonly referred to as "the brain's wiring diagram". For simplicity, we shall mainly focus on structural connectomics at the microscale, though many of the points raised will apply to the other scales as well. At this scale, the complete connectome is a map of all the synaptic connections between all of the neurons in the system at hand, both electrical and chemical. Whether it is directed or not is a matter of convention. There are no non-invasive techniques which allows this level of detail to be determined, so the analysis can only be done post-mortem on thin slices of brain tissue. This makes it hard to study the relationship between the structural connectome and higher-level cognitive features, since this naturally requires a living organism. In practice, by staining the tissue in the thin slices of the brain, the circuit diagram is obtained by visualising the synaptic vesicles in serial brain sections using an electron microscope. Using this approach, the direction of the connections can be determined, but it is usually unknown whether the synapses are excitatory or inhibitory. As of now, full connectomes have only been determined for a few small model organisms.

The connectome of the worm *C. Elegans*, a commonly used model organism in biology, was found in 1986, after 10 years of effort. This consists of only 302 neurons with about 7000 chemical synapses, and showed a surprising level of interconnectedness and complexity, making the system challenging to analyse [11]. Much simpler and easier to analyse is the connectivity diagram of the *stomatogastric ganglion* (STG) of the crab *Cancer borealis*, consisting of only about 27 neurons [4]. There is also an ongoing effort to map the full connectome of the *Drosophila* fruit fly, which has about 135.000 neurons. Far more ambitious is the Human Connectome Project, which aims at obtaining a complete map of the human connectome [17].

The main issue, and a commonly raised concern regarding the usefulness and explanatory power of such structural diagrams, is that they are static descriptions of dynamical systems

system, taken at a specific instant in time. The overarching aim of much of this research is to understand how complex functions, such as behaviour, arise from processes at lower levels, such as those between neurons. These include the firing of action potentials, the release and uptake of neurotransmitters at chemical synapses, as well as the release of neuromodulators and neuropeptides, amongst others. Importantly, the information processing at this level, from which the higher order functions are thought to emerge, are *dynamic* processes which cannot be described purely in terms of static connections. Yet they are by no means independent of the structural connections, so the aim of structural connectomics is by and large to find the mapping between the static wiring diagram and the higher-level functions. The question then is to what extent these dynamical processes are indeed reducible to the static structure, and much has been written about the limitations and potential challenges related to this approach [38]. These criticisms generally fall under one of two categories. The first is related to the plasticity of the brain at different timescales, the second to the sheer complexity of even the static diagram and the ensuing interpretational issues.

The brain's plasticity is a key feature of its functioning. It is known to be plastic in numerous ways and at different timescales, meaning that the brain is constantly changing in ways which affect the information processing and the functional properties of the circuits involved. These changes, though, are not reflected in the structural connectome. At short timescales, the main sources of plasticity (known as short-term plasticity) are neuromodulators and neuropeptides. Neuromodulators are chemical substances which do not directly affect the ion-channel receptors, but which nevertheless modulate the function, activity and communication between the neurons by regulating the response of the receptors at the post-synaptic site. Examples include dopamine, noradrenaline and serotonin. Similarly, neuropeptides are small chains of amino acids which modulate neural activity in numerous ways. Examples include endorphins and oxytocin. Both neuromodulators and neuropeptides are often released near synapses but are not necessarily passed on to a single neuron in the same way as the neurotransmitters released at chemical synapses. Instead, they can diffuse through greater areas in the brain, in a manner similar to that of hormones, through the circulatory system. This makes their pathways impossible to detect through electron microscopy, and they therefore lack a structural correlate which could form part of the structural connectome. Research by Marder and Bargmann has shown that both neuromodulators and neuropeptides have the potential to greatly affect the function of neural circuits in a range of different ways [4]. In the case of the STG, the effective circuit diagram at any moment is largely determined by the current neuromodulatory environment. For example, it has been found that synaptic connections can be functionally turned on or off, entire circuits silenced, and the spiking patterns of individual neurons changed [35]. In *Drosophila*, being in a state of starvation has been shown to increase the sensitivity of the fly's behavioural response to sugar, which is in turn partly dependent on the neuromodulator dopamine [26]. Similar effects have been found in mammals. This shows how the anatomical connectivity on its own is insufficient for determining the function of neural circuits. Instead, it can be thought of as a map of the set of *potential* functional circuits which a system can instantiate given different neuromodulatory environments. To get an idea of the extent of the underdetermination of the wiring diagram in determining circulatory function, consider that although *C. Elegans* only has 302 neurons, its genome encodes over 200 different neuropeptides, suggesting an immense potential for modulation of the connectome, giving a multitude of different functional circuits for

the same structure [56].

This form of neural plasticity creates a further problem in determining the mapping between structure and function. Neuromodulators and neuropeptides can spread without anatomical synapses to areas both far and near, and at short timescales, making it impossible to precisely measure the modulatory state over time. Indeed, neurobiologist Cornelia Bargmann writes that "there is no reliable way to assess the complete modulatory state within any animal, including *C. elegans* – it is the dark energy of the nervous system, inferred but not measured." [3, p. 462]. This introduces a substantial problem when trying to map structure to function, with a danger of circular reasoning, since changes in the dependent variable (the functional properties) are used to infer the presence of modulatory factors. The circularity can be summarised as follows:

1. We are trying to determine the mapping from structure *(S)* to function *(F)*, call this $F = m(S)$. However, this relation is known not to be true, as the function is also affected by the environment, call this E. We thus have the relation $F = m(S, E)$.

2. F and S are the only observable variables, so we are trying to determine both the function m and the environment E from these.

3. Neither E nor m can be tested empirically, since that would require knowing the other three variables.

4. Thus, we are trying to determine both the function *and* one of its independent variables solely from the output and the other independent variable, i.e., using the function m to infer E from F and S whilst simultaneously using the variable E to infer m from F and S.

Thus, almost any function can be explained by assuming a new modulatory state, and vice versa, making it very difficult to extract the contribution of the structure to the function. This greatly constrains the explanatory power of the structural connectivity diagram beyond providing a vast number of potential functions a circuit could perform in different (largely unmeasurable) environments, and even these will be difficult to test empirically.

At longer timescales, a significant source of change in the wiring diagram is the synaptic rewiring which takes place during development and learning. Such changes are more trackable, but it nevertheless means that the connectome cannot be treated as a stable description of even the "static" wiring diagram of an organism. Thus, depending on the degree of plasticity, it will be of limited explanatory value as time passes.

Another problem which is sometimes raised regarding the mapping between function and structure is that in addition to the functional ambiguity of circuits, functions are also multiple realisable by different circuit motifs, meaning that the same function can arise from different circuit structures [38]. However, assuming that the main purpose of the mapping is to be able to infer function from structure, and not the other way around, a one-to-many mapping from function to structure is unproblematic, apart from the additional complexity it introduces into the data analysis.

Finally, there is the problem of the sheer complexity of the system. Even if one assumed that all of the activity of the network was mediated through the synaptic connections described in the structural connectome, interpreting this has proven very challenging. The

connectomes studied so far have been found to be highly interconnected, and it is not uncommon to find seemingly "contradictory" motifs in the same connectome. For example, in the STG, there are many instances of parallel pathways between the same two neurons, sometimes with opposite effects, neurons connected through reciprocal inhibition or synapses which are dormant [11]. In the case of *C. Elegans*, it is usually possible to find a path between any two neurons in only three synapses, so to assign functional roles to edges and motifs has proven far harder than determining the functions of individual neurons. Regarding this, Bargmann and Marder write:

> At this point, over 60 % of *C. Elegans* neuron types have defined functions in one or more behaviours. This notable success, however, hides a surprising failure. For *C. Elegans*, although we know what most of the neurons do, we do not know what most of the connections do, we do not know which chemical connections are excitatory or inhibitory, and we cannot easily predict which connections will be important from the wiring diagram [4, p. 485].

Thus, the connectome also lacks the information about the relative significance of different synapses and their actual contributions to the functioning of the system.

To summarise, the structural connectome of the brain is a static map of the synaptic connectivity of the brain, and must be treated as such. It is very far from a complete description of the dynamic information processing which takes place in the brain and from which higher-level functional properties arise. However, even though these high-level properties are not fully reducible to the anatomical structure, the latter certainly plays an essential role in constraining the potential functional circuits which can be instantiated in the brain.

Functional Connectomics Whereas the structural connectome shows the anatomical connections in the brain, the functional connectome is a description of the *statistical correlations* between activity in the different units of interest. Often, one also try to correlate this activity with higher-level cognitive functions. As with the other types of connectomics, this can be done at multiple scales. In animals, where in-vivo invasive methods are permissible, one can insert electrodes into the brain to record from individual neurons or smaller clusters of neurons while also recording the activity of the animal. With humans, it is most often done at the macro-scale, involving recordings of different types of activity measured at the surface of the skull. Since this can be done without non-invasively, it is the most commonly used connectivity analysis method in humans. Methods include, amongst others, magnetoencephalography (MEG), electroencephalography (EEG) and functional magnetic resonance imaging (fMRI), of which fMRI seems to be the most common method in functional connectivity studies, so shall be the primary focus here. One reason for this is that the nodes derived from the fMRI data are voxels (a volume-based unit which the brain can be divided into), i.e. anatomically localised regions in the brain. The corresponding nodes in EEG and MEG are the recording electrodes or the sensors on the surface of the skull, which are less suitable when the aim is to determine a three-dimensional anatomical functional connectome of the brain.

fMRI has a spatial resolution on the order of millimetres and a temporal resolution on the scale of second. This is because fMRI measures the activity indirectly by monitoring

the level of oxygenation in the blood, or more specifically, the iron atoms in oxygenated and deoxygenated haemoglobin. The ensuing signal is known as the blood oxygen level dependent (BOLD) signal. This is related to neural activity because activity in neurons leads them to consume oxygen, which involves their deoxygenation of the haemoglobin. The connection has been experimentally verified, so the BOLD signal is to some extent a good proxy for neural activity, though it comes with a certain time-lag and low temporal resolution [12]. An underlying assumption in most fMRI studies is that of functional localisation, meaning that higher-order functions are localised to smaller regions of the brain. There is evidence showing that this if often the case. For example, the high degrees of clustering in the brain ensures that neurons which are spatially close together tend to fire together and process the same information [51]. However, the exact degree of localisation differs from function to function, so is not something which one can simply assume *a priori*. This therefore limits the range of functions which can be studied using fMRI.

The standard setup in fMRI studies is to place the subjects in an MRI scanner and record changes in the BOLD signal as the subjects perform certain tasks as instructed. The resultant signal is then usually compared with the signal recorded when the subject is in a "resting state", meaning that they are just sitting there without performing any cognitive task. This resting state activity is known as the *default mode network*. The idea is then that the difference between the activity maps in the resting state and the task-related state can be linked to the functions required to perform the task at hand. This method is known as cognitive subtraction [60]. Often, this is repeated for numerous related tasks, with the hope that some of the same regions will light up for different tasks, such that these can be functionally related to what we take as the commonalities between the tasks. Thus, by combining assumptions about functional localisation and connectivity, one might hope to determine how different functions in the brain relate and support each other.

The main problem here, however, is not with the inferred connectivity diagrams themselves, but rather with how they are used. As mentioned, such studies are often conducted with the aim of not just determining the correlations between activity in different regions of the brain, but comes with an additional premise about the higher-order function these regions perform. The functional connectivity then becomes a map of the interdependence of different functions of the brain. This is a rather controversial endeavour, since there is much debate about the localisation of function in the brain and how to infer it [34].

First of all, the increase in activity in an area tells us little about the actual function performed in that region, and might even just be a spurious correlation between location and function. All we can really tell from such activity is that the voxels in question are probably contributing with something which is necessary to perform the task at hand. However, we do not know how the brain decomposes a cognitive task into its logical parts, so even though different tasks might induce activity in the same regions, it is difficult to know what logical part of the task this corresponds to, or the significance of the contribution. Furthermore, the same task will usually lead to increased activity in multiple regions, making it even harder to connect them to specific functions. Conversely, most brain regions are activated in a lot of different functional contexts.

Further, there is the danger of picking up random fluctuations in the activity. This point has rarely been made as succinctly as in an fMRI study conducted by Bennett and colleagues. As they describe it:

> One mature Atlantic Salmon (*Salmo salar*) participated in the fMRI study. The salmon was approximately 18 inches long, weighed 3.8 lbs, and was not alive at the time of scanning. [...] The salmon was shown a series of photographs depicting individuals in social situations with a specified emotional valence. The salmon was asked to determine what emotion the individual in the photo must have been experiencing [5].

Activity was detected in multiple voxels within a cluster spanning 81 mm^3, with a statistical significance of $p = 0.001$, which, by the standards of the field, is sufficient to conclude that the salmon was indeed processing the images in the specified regions. Although a striking example of the dangers of blindly trusting fMRI results, it should be noted that such results are avoidable by simply repeating the same experiment or by taking the average over a larger sample of participants in the studies.

A final concern regarding functional connectomics is the resolution of the fMRI images and the functional scale this corresponds to in the brain. Looking at the 300 top-cited cognitive fMRI studies as of 2008, neuroscientist Nikos Logothetis found that the average voxel size studied was approximately 55 mm^3 in size. He further estimated that such a volume of brain tissue contains around 5.5 million neurons, between 2.2 and 5.5×10^{10} synapses, 22 km of dendrites and 220 km of axons [34]. This stands in a stark contrast to the scale at which the structural connectome is usually studied, where function is typically assigned at the level of individual synapses and circuit motifs. Whereas it is possible to analytically determine the theoretical potential for the information processing of smaller circuit motifs (see for example [1]), it is as yet unknown how this scales when combined in such grandiose numbers. The analysis of fMRI studies are thus further complicated by the fact that we do not know what the potential functional extent of individual voxels might be. Further, this resolution also imposes a limit on what activity can be detected, since scant or spatially distributed (but potentially functionally significant) activity is largely undetectable using fMRI, making the functional connectivity map an incomplete description of the actual activity.

In light of these concerns, philosopher of neuroscience Carl Craver has argued that the main uses of fMRI studies are system identification, brain parcellation and comparative analysis [13]. System identification involves detecting clusters of highly connected neurons by looking at voxels of neurons which tend to fire together. This is based on the Hebbian idea that "neurons that fire together wire together", which in turn should cause their activity to become related. The assumption is that these also form functional units, but the analysis itself can be done without attributing any function to these clusters. Conversely, the same methods of analysis can be used to find discontinuities in the activity, which can be indicative of cortical boundaries where the neurons on each side of the discontinuity are unconnected. This might then be indicative of a functional boundary. Comparative studies can be used to compare brains of different individuals looking for systematic differences. This can in turn be used for diagnostic purposes of neural disorders or illnesses, where statistically significant differences between populations of healthy and unhealthy patients can be used either as an indication of where the problem might reside or simply as a correlation-based diagnostic tool. It should be noted, though, that Craver stresses that the fMRI studies in all of these cases are insufficient for providing proofs or explanations of the suggested findings, and that

determining their correctness requires further verification using other methods.

Effective Connectomics Finally, we have effective connectivity. This corresponds to a connectivity diagram delineating the causal connections between the nodes of the network. Comparing it to the structural connectome, it would only include the directed edges (chemical synapses) which are functionally active. It can either be taken as a time-average of the causal influence between the neurons, or as a time-varying description of these. For a complete description of all the causally effective components of the system, it would also include the source- and target-neurons of neuromodulators and neuropeptides, though this is difficult as a result of the aforementioned challenges in measuring the neuromodulatory environment.

One of the major challenges in determining the causal connectivity of neural networks is the sheer size of the system. At the microscale, it is possible to insert electrodes into the brain which can detect the spiking patterns of individual neurons. One could imagine that given a complete description of the activity of neurons in a fairly segregated cluster, it would be possible to infer the causal connectivity between the neurons given enough data, both observational and interventional. However, the number of connections makes this a surprisingly difficult endeavour. To illustrate, a group of researchers at the Max Planck Institute set out to reconstruct a tiny bit of the wiring in the mouse brain [40]. They looked at a cubic part with sides less than 0.1 mm long, containing 89 neurons. Yet, in this tiny structure they found a staggering 153,171 synapses. This makes the dynamics of the system very complex, with a potential for a range of different stable patterns of activity. Thus, if only given the spike trains of the neurons in question, it will be a difficult endeavour to establish the effective connections between all the pairs of neurons. Further, given the change in the causal dependencies over time, there simply is no "one true" effective connectome.

This example of the mouse brain also illustrates another problem with causal inference, which is that of confounding variables. In the case above, the tiny fraction studied not only contained immense numbers of internal connections. The 0.001 m^3 of the brain contained 2.7m of axons and dendrites whose cell bodies were located outside of the volume studied, exerting causal effects whose sources will not be included in the dataset. In the lingo of causal inference, such unmeasured but causally effective variables are known as *confounding variables*. This makes determining the causal structure between the elements observed very difficult, since it is often impossible to differentiate between direct causal effects, causal effects mediated by other neurons and effects which have a common cause. Given the size and complexity of the brain, the causal diagram will usually be largely underdetermined.

Finally, even if such a complete effective connectome could be obtained, it is far from clear how a human would use it to gain a deeper understanding of the system, given its size and complexity. In terms of Marr's levels of explanation, although we are able to find the implementational and algorithmic explanations of the lower level dynamics, connecting these to the purpose of the computations requires a more holistic understanding of the system and how the different levels supervene on each other. This interpretational issue also pertains to the structural and functional connectomes, and shows the need for explanatory models and bridge principles connecting the different levels of interpretation of the brain.

2.1.3 The Need for Models and Idealisations

Following from this, it might seem a hopeless cause to try to understand how the brain works. And indeed, there is general agreement that the brain is amongst the most complex systems science is grappling with at the moment, and that we are very far form understanding it. What is needed are theories and simplifying models for moving from the vast amounts of raw data to both a holistic understanding of the system as a whole, as well as a conceptual understanding of the underlying principles and how these are connected[59]. This is not a new problem, and all branches of science operates with a range of different simplifying models to obtain this. These are partial descriptions of phenomena that usually focus on some aspects of them whilst deliberately disregarding others. In that sense, they always contain some known falsehoods, yet are nevertheless seen as necessary for providing intuition and understanding. This might seem paradoxical, but there are many advocates of the thesis that truth and understanding do not necessarily come hand in hand. Philosopher Nancy Cartwright makes this point in the opening lines of her essay, poignantly entitled *The Truth Doesn't Explain Much*: "Scientific theories must tell us both what is true in nature, and how we are to explain it. I shall argue that these are entirely different functions and should be kept distinct. Usually the two are conflated" [10, p. 44]. The reason for this is that the complete truth about a system is usually too complex for humans to understand in its full detail. In another essay, published in the same collection, she claims that to "explain a phenomenon is to find a model that fits it into the basic framework of the theory and that allows us to derive analogues for the messy and complicated phenomenological laws which are true of it" [10, p.152]. Models are thus necessary to explain a phenomenon, and they do so either by working as a mediator between the theory and the phenomenon or by representing the phenomenon directly, independent of a theory.

Regarding the known falsehoods, such as variables which get removed from the model description although they are known to influence the variables of interest, Catherine Elgin presents a compelling argument in favour of such idealisations [18]. According to her, models represent phenomena by exemplifying certain features or properties that then gets imputed onto the system in question. These are features that might otherwise get overlooked, so by exemplifying and emphasising these, models give epistemic access to significant properties that might be hard to discern otherwise. Furthermore, such exemplification is not necessarily best done by conveying a mirror-image of the system. Indeed, one of the main functions of scientific models is to disentangle the important properties of a complex system from the more arbitrary, negligible and irrelevant ones. It can therefore be useful to include what she calls "felicitous falsehoods", such as idealisations, simplifications, distortions or abstractions, in the model in order to make the important features of the phenomenon more salient.

Thus, if the hope is that humans should not just be able to use computers to make predictions about the brain, but also understand some of its underlying mechanisms and causal dependencies, we need some simplifying models which reduce the number of components and their types dramatically. This allows us to study how the variables depend on each other and how changing some of these affect both their mutual dynamics and properties of the system as a whole. This was the approach taken in this project, where all the data analysed was computationally generated using a neural dynamics simulator, which in turn was based on a range of simplifying models. These models, along with a detailed description of the

exact model used to generate the simulated data studied in this project, are presented in section 2.5.

2.2 Simplicial Complexes

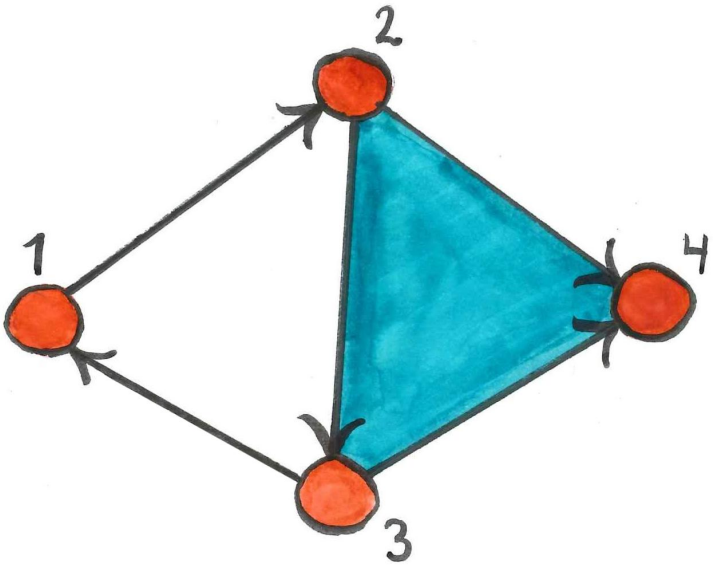


Figure 3: A directed graph consisting of four nodes. Nodes 1-2-3 form a cycle, so is not considered a directed simplex. Nodes 2-3-4, however, form a directed 2-simplex with 2 as the source, 3 as the mediator, 4 as the sink. Thus, 2 is the source of a 2-simplex, even though it has an incoming edge from node 1, since node 1 is not a part of said simplex.

The primary aim of this project was to analyse the functional role of simplicial structures in biological neural networks, so an introduction to simplicial complexes is required. Simplicial complexes were first introduced in algebraic topology as a generalisation of graphs, moving from a focus on dyadic relationships between two and two elements to also include higher-order relationships as units of interest. Following the notation of [21], a directed graph $\mathcal{G}$ can be represented as a set of vertices and edges (V, E), and a function $\tau = (\tau_1, \tau_2)$. V denotes the set of vertices, $\{v_i\}$, and E the set of edges, $\{e_i\}$, of $\mathcal{G}$, and the function τ specifies the direction and vertices of each edge. Specifically, if the edge e goes from vertex v_1 to v_2, we have that $\tau(e) = (v_1, v_2)$. Further, the functions τ_1 and τ_2 returns the source and target vertices of the edge respectively, such that $\tau_1(e) = v_1$ and $\tau_2(e) = v_2$. We also require that there are no self-loops and that there is at most one edge with the same direction between each pair of nodes. Note that this still allows for reciprocal connections between vertices.

This framework gives a representation of a graph which is centred around the dyadic relationships between pairs of vertices. There are many other types of relationships which may exist between nodes in a graph, however, and one such type of higher-order relationship which has recently become a popular target of investigation is *triadic* relationships between three and three elements. This is the inspiration for the notion of a *simplicial complex*.

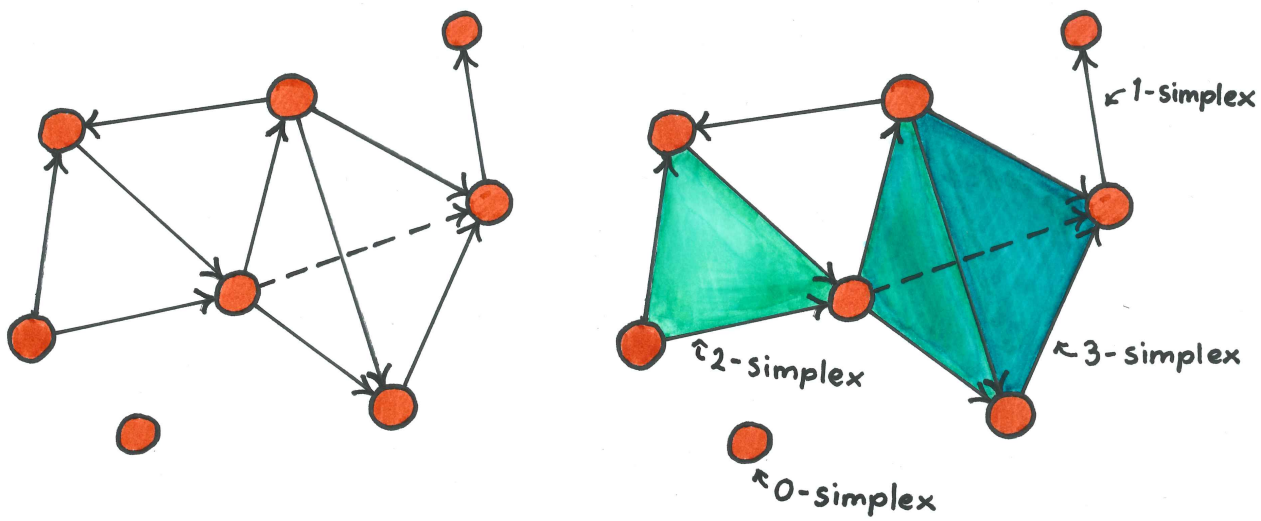


Figure 4: A directed graph (left), with its corresponding simplicial complex to the right. Simplices of different dimensions are labelled as examples.

In the same way as a graph, $\mathcal{G}$, can be represented by its vertices and the edges between them, an abstract simplicial complex, $\mathcal{S}$ consists of a set of vertices, as well as sets of *simplices* of different dimensions. Originally, such simplices were restricted to undirected graphs. A 2-simplex then consists of three vertices which are all connected to each other, it exists in two dimensions and is commonly called a triangle. Similarly, a 3-simplex consists of four interconnected vertices, is 3-dimensional, and is known as a tetrahedron. In general, an *n-simplex* spans n dimensions and consists of $(n+1)$ vertices which are all directly connected to all the other vertices in the simplex. For consistency, a vertex is denoted as a 0-simplex, and two vertices connected by a single edge is a 1-simplex. We denote each simplex σ, regardless of its dimensionality, such that the simplicial complex $\mathcal{S}$ consists of the set of all the simplices $\{\sigma_i\}$. Each simplex σ is itself a set containing the constituent vertices of the simplex. The dimension of each σ is its cardinality minus one, such that if σ consists of $(n+1)$ vertices, it is an *n-simplex* of $\mathcal{S}$. It is required that each subset σ_j of σ is also in $\mathcal{S}$. The set of $(n\text{-}1)$-dimensional simplices which forms part of the n-dimensional σ are called the *faces* of σ. Thus, for each $\sigma = (v_0, v_1, ..., v_n) \in \mathcal{S}$, for each $0 \leq i \leq n$, the i^{th} face of σ is an $(n\text{-}1)$-simplex obtained by removing the vertex v_{n-i}. Because of the requirement that every face of a simplex is also a simplex, it is sufficient to specify only the set of maximal simplices, which are those that do not appear as the face of any other simplex. The set of all the maximal simplices is thus sufficient to determine the entire simplicial complex, since every simplex is either maximal itself, or the face of some other maximal simplex.

This framework for analysing graphs in terms of simplices can be generalised to directed graphs, as presented in [47]. In the case of a directed graph, an *abstract directed simplicial complex* is needed. Here, the elements σ of $\mathcal{S}$ are ordered sets of vertices, where the order determines the direction of the edges. Further, only simplices with a single source and a single sink are defined as directed simplices. In this case, the source is the vertex with only outgoing edges in the given simplex, and the sink the vertex with only incoming

edges. This requirement holds for simplices of any dimension, such that an n-dimensional simplex, consisting of $(n+1)$ vertices, will have one source-vertex, one sink-vertex, and $(n\text{-}1)$ mediator-vertices. Note that a vertex can be a source/sink and still have incoming/outgoing edges, as long as these do not form part of the same simplex. See fig. 3 for an example of this requirement of directionality, and fig. 4 as an example of a directed graph with its corresponding simplicial complex.

Each simplex can be represented as a list of nodes, where their order gives the structure of the simplex. For each node, there is an outgoing node from this to all the subsequent nodes in the list. Thus, for an n-simplex, the first node has n outgoing edges, the second has $(n\text{-}1)$ outgoing and 1 incoming edge, the third has $(n\text{-}2)$ incoming and 2 outgoing edges and so on. Further, every subset of the list of nodes which respects the order of the maximal simplex will be a face of the maximal simplex. For example, the 4-simplex $[a, b, c, d, e]$ is made up of this set of 3-simplices $\{[b, c, d, e], [a, c, d, e], [a, b, c, e], [a, b, c, d], [a, b, c, d]\}$ and this set of 2-simplices $\{[a, b, c], [a, b, d], [a, b, e], [a, c, d], [a, c, e], [a, d, e], [b, c, d], [b, c, e], [b, d, e], [c, d, e]\}$.

2.2.1 Hasse Diagrams

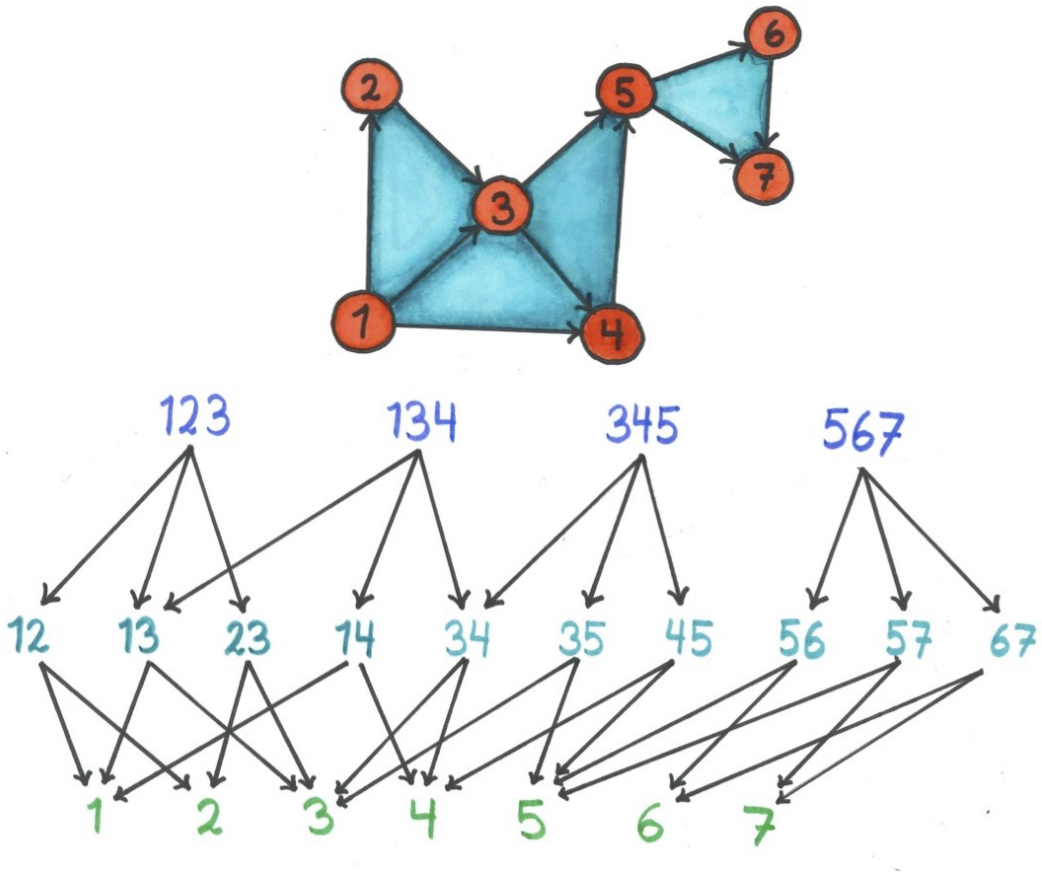


Figure 5: Diagram of a directed simplicial complex (top) and its corresponding Hasse diagram (bottom).

A Hasse diagram is a mathematical diagram which can be used to represent ordered

sets, such as simplicial complexes. The entire binary (ignoring node properties and the edge weights) representation of the connectivity graph can be represented in a single Hasse diagram. This is ordered hierarchically, where each level of the Hasse diagram contains simplices of the corresponding dimensionality. Assuming a simplicial complex $\mathcal{S}$, the level n vertices of the corresponding Hasse diagram $\mathcal{H}$ are the n-simplices of $\mathcal{S}$. The 0^{th} level are the nodes, i.e. the 0-simplices, the 1^{st} level are the edges, i.e. the 1-simplices, the 2^{nd} level the 2-simplices etc. Further, the vertices at each level of $\mathcal{H}$ are oriented from left to right corresponding to their position in the simplex. An example of a simplicial complex and its corresponding Hasse diagram is shown in fig. 5. Importantly, all the information about the "descendant" lower levels of a simplex is contained in the vertex at the highest level.

In [47], they present an algorithm for extracting the Hasse diagram of a regular directed graph in matrix form. The algorithm is presented in their supplementary material. As part of the current project, we implemented this algorithm from scratch in Python in two variants, using `numpy` and `PyTorch`, since the original implementation was in C++.

2.2.2 Relevance for Neuroscience

Given the interpretational and computational challenges discussed in section 2.1.2 above, there is an increasing interest in the fields of neuroscience and connectomics in finding alternative methods for analysing and understanding neural networks [51]. A common approach is to try to understand the structure and computational potential of neural networks in terms of the basic units of computation and the flow of information between them. Traditionally, this has meant looking at individual neurons, clusters or regions, where dyads (i.e. two units connected to each other) has been treated as the basic units of the flow of information. These can then be combined to form greater network motifs. However, this approach assumes that the features of interest at higher levels are reducible to these low-level dyadic relationships. This is not necessarily the case, as many argue for the existence of emergent properties in complex systems, including the brain [51]. Such emergent properties are observable when one looks at the system at certain levels, without being reducible to the lower-level components of the system. Thus, if we only study the system at those lower levels, we are simply unable to explain the emergent features, because we are looking at it at the wrong level. This would be akin to trying to use particle physics to explain the pollination of roses. An obvious challenge in a new and emerging field, such as network neuroscience, is that one does not know *a priori* what the correct level of analysis is for any phenomenon. In other words, assuming some form of emergence, it is impossible to know in advance what the lowest possible explanatory level is for the phenomenon one wishes to explain. As for any research programme in its infancy, this thus comes with its due amount of speculation, trial and error.

In the case of connectomics, this focus on dyadic relationships might lead us to miss out on important higher-order structures which are causally relevant for the emerging functionality of the brain at different levels. Hence, there is a growing interest in exploring the potential for analysing biological neural networks in terms of other higher-order network structures, such as their simplicial structure. The motivation for this is natural. It is well known that clusters of highly connected neurons tend to fire together, so structural connectivity is tightly linked with functional connectivity [25]. In the lingo of motif analysis, a simplex of a

given dimension is a motif, since all simplices of the same dimension have the same number of nodes and edges. Further, given the strict constraints on the directionality of directed simplices, all directed simplices of the same dimension are isomorphic, meaning that only one motif class of each graph motif is included in the set of directed simplices.

In their [21], Giusti, Ghrist and Bassett present an overview of current research into analysing the functional connectivity in terms of simplices rather than dyads. As an example, simplicial complexes can be used to represent higher order correlations between the firing rates, such as cofiring over time. Strong coactivity between neurons can be a sign that these are connected and somehow form a functional unit which processes related pieces of information. The fact that this focuses on functional connectivity rather than structural connectivity means that in experiments, where the sensory input can be controlled, it is possible to connect certain sensory input to activity in particular simplicial structures. As an example, Curto and Itzkov [14] managed to use such higher-order patterns of coactivity in hippocampal place cells to extract global features of the animal's environment, as well as a reconstruction of the topological map of the environment. Multiple studies have also applied some sort of weight filtering to these coactivity complexes. This involves using a continuous filter on the complex, where only those edges with weights (correlations) above a certain threshold are kept. This approach has successfully been used to distinguish between subjects with different neurological and cognitive difficulties and control subjects, including classification of children with attention deficit hyperactivity (ADHD), autism spectrum disorder (ASD) and control subjects [31] and differentiating between pre- and postlingual deaf adults [29].

Another approach is to focus on simplices in the *structural* connectome. Here the nodes represent the neurons and the edges the chemical synapses between them. This can be done both with and without regard for the direction of the synapses, but in the case where their direction is taken into consideration, the information flow is uni-directional along any given edge. The bidirectional electrical synapses are thus ignored. Higher-order structures of neural networks might then be informative of both local and global properties of the information flow in the network. One way of analysing such networks is thus by representing them as a directed simplicial complex, where each simplex has a single source neuron and a single sink neuron. This is to prevent cyclic relationships from being included, ensuring that the causal flow is uni-directional between the source and the sink.

This was the approach taken by Reimann et al. [47], whose aim it was to study the relationship between the structural and functional simplicial complexes. They analysed digital reconstructions of rat neocortical microcircuitry, resembling biological neural networks in terms of density of neurons and connectivity patterns. The functional data came from simulating activity in these networks. The functional connectome was taken to be a subset of the structural connectome, containing only those synapses which were active within a certain time-frame (i.e. those where the post-synaptic neuron fired shortly after the pre-synaptic neuron). Thus, the functional graph represented the pairs of neurons which fired subsequently within a given time-frame, and *not* the simultaneous co-activity, as in [21].

They compared the simplicial structure of the reconstructed neocortical circuit graphs with five Erdös-Rényi random graphs with similar graph properties, such as size and connectivity. They found that the number of higher-dimensional simplices in the biological networks far exceeded those in the random graphs, and this discrepancy increased with di-

mension. Performing the same analysis on the structural connectome of *C. Elegans* gave the same result. This does not in and of itself prove anything about the usefulness or functional significance of such higher-order simplices, as they could simply be a byproduct of something else. However, the fact that they are found in unexpectedly high numbers across species is at least an indication that they are possibly performing some significant functional role which has been selected for through evolution.

Regarding the functional analysis, their main approach was to study the spike correlations between different neurons as function of their simplicial role. This was done for microcircuit activity in response to sensory stimuli in activity simulations in the reconstructed networks. Comparing the pairwise correlations between all pairs of neurons in these networks, they found that this increased significantly with the number and dimension of the directed simplices to which the neurons belonged. They took this correlation, which implies a form of coactivity, as a measure of emergent integrated activity patterns, which is in turn believed to be indicative of more complex patterns of activity and information processing. They took this finding to mean that "the hierarchical structure of the network shapes a hierarchy of correlated activity" [47, p. 2]. Similarly, they established that the functional correlations between neurons in the same simplex depended on their role therein. They found that the spike correlation between any pair of neurons increases with the pair's proximity to the sink of the simplex, arguing that this results from the corresponding increase in shared informational input. From these observations, they conclude that "the emergence of correlated activity mirrors the topological complexity of the network." [47, p. 10]

One problem with this approach is the strict emphasis on *correlations*. As discussed in the section on functional connectomics, correlations in activity are insufficient for establishing causation, which is generally seen as a better method for studying the ontological structure of systems. Indeed, descriptions of correlations can hardly be said to be explanatory, as most agree that scientific explanations are required to be causal. Thus, if we want an explanation of the neural dynamics, we need to establish the causal relationships between the network's constituent elements. This project was largely based on the motivations and findings of [47] and [21], but the emphasis was on the causal relationships rather than the correlational ones. The initial idea was to test whether the directed simplices could potentially be treated as separable causal units in the network. Assuming this is the case, it is reasonable to assume that their causal and functional role will be instantiated and thereby observable at different scales and using different metrics. Further motivation, as well as how this was approached and tested, will be further elaborated below.

2.3 Causal Inference

In very broad terms, the overarching aim of most scientific endeavours is to establish the underlying causal structure of the system under investigation. Somewhat surprisingly, then, a rigorous mathematical framework for how to extract such causal information is fairly new, and has become known as *causal inference*. The primary aim of the field is to establish what empirical evidence is required to infer causal relationships and what legitimate inferences one might indeed draw given different types of information in different situations. In his [45], Judea Pearl presents this framework, on which the following exposition is largely based.

Although statistics and probabilities can tell you about the statistical dependencies be-

tween variables, without further specification, it is based on purely correlational measures. Pearl argues that the lack of information about the origins and mechanisms giving rise to the correlations makes such correlation-based models less stable and robust than those based on causality. He writes: "We expect such difference in stability because causal relationships are *ontological*, describing objective physical constrains in our world, whereas probabilistic relationships are *epistemic*, reflecting what we know or believe about the world. Therefore, causal relationships should remain unaltered as long as no change has taken place in the environment, even when our knowledge about the environment undergoes changes." [45, p. 25] Thus, if the aim is to truly understand a dynamical system, knowing the causal relationships between the variables will likely be of far more value than knowing their correlations.

Structural Causal Models The main property of causality is in many ways its asymmetry: causes precede their effects. Nevertheless, when we express quantifiable relations mathematically, we tend to do so using functions with equalities, even when the relations are causal. The problem with this is that the same equation can be rewritten in many ways, and in that sense only express the *correlations* between the variables. Take for example Newton's second law, stating that $F = ma$. This presents the relations in a symmetric way, where any one of the variables can be expressed as a function of the other two, such that, given any two of the quantities, the third can be found. Considering the causal structure of these variables, however, we would say that the force causes the acceleration, not the other way around, so by expressing it in an asymmetric fashion, we leave out essential information about the system. Instead, we could express it using a *structural equation*, which also captures the causal structure of the system. In this case, we would rewrite it as

$$a := f(F, m) = \frac{F}{m},$$

where the expression $:=$ signifies that the acceleration a is the effect, and the independent variables of the function f the causes, so it encodes the causal direction and asymmetry of the situation. In general, this gives rise to the notion of a *structural causal model*. Such models specify all the causal dependencies between the variables in the system, and can be expressed either as a set of structural equations, or as graphs.

We use uppercase letters to denote the variables in the model, and lowercase letters to denote the values which they can take. Thus, the variable X may take on any of the values from the set $\{x_1, x_2, ..., x_n\}$. The structural equation of X is a function whose independent variables are the direct causes of the X. In addition, there will usually be some noise or background conditions which are not accounted for in the model, but which will nevertheless affect the variable X. These unknown background conditions are grouped together in the random variable U, which introduces some uncertainty in the outcome. An example of such a structural causal model, M, is shown in fig. 6. The structural equations for the same model M are

$$\begin{aligned} B &:= f_B(A, U_B) \\ C &:= f_C(A, B, U_C) \\ D &:= f_D(A, C, U_D) \end{aligned} \tag{1}$$

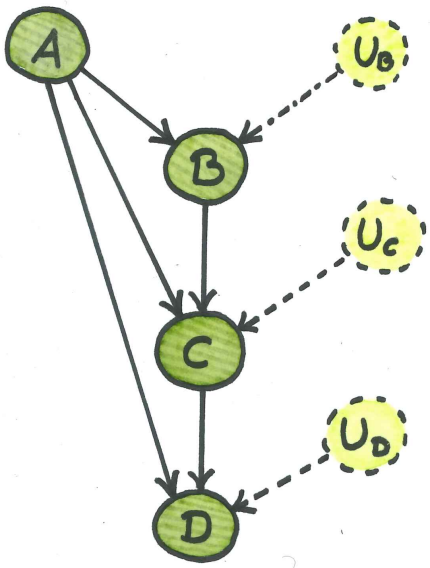


Figure 6: A structural causal model corresponding to the structural equations in eq. (1). The variables *A, B, C* and *D* are the observed variables, whereas the variables in the dashed circles are their respective noise variables, which remain unobserved.

For adjacent nodes A and B, where there is an edge from A to B, A is referred to as the *parent* of B and B as the *child* of A. Only the direct causes, i.e. the parents, of a variable are included in its structural equation. If there is a path from variable A to C which passes through B, without there being any direct edge from A to C, then A is an *ancestor* of C, and C a *descendant* of A.

It is usually also required that the model is a *directed acyclic graph* (DAG), meaning that (within a given time-step) there are no reciprocal edges between two nodes and there cannot be a directed path from any one node to itself. It is then possible to use such a model to find the effect which intervening on one of the variables has on the other variables. This is something which a purely correlational model cannot provide, since the asymmetry of causation is missing from this description.

Causality is closely linked with the notions of *intervention* and *counterfactuals*. An intervention is an action where one intervenes by assigning a value to one of the variables in a causal model without affecting the others. This will affect the children, and in turn the descendants, of this variable. Given a variable X, we denote the intervention of assigning value x to X as *do(X=x)*. By looking at the changes in the probability distributions of the descendants, it is then possible to extract the causal effects of variable X on the other variables in the model. This gives rise to Pearl's definition of causal effect:

Definition 2.1: Causal Effect

"Given two disjoint sets of variables, X and Y, the causal effect of X on Y, denoted either as $P(y|\hat{x})$ or as $P(y|do(x))$, is a function from X to the space of probability distributions on Y. For each realisation x of X, $P(y|\hat{x})$ gives the probability of $Y = y$ induced by deleting from the structural equations of X all equations corresponding to variables in X and substituting $X = x$ in the remaining equations." [45, p. 70]

Confounding Variables A considerable challenge when inferring causal relations are confounding variables. These are unobserved variables which might influence the observed variables in significant ways, which might lead us to make incorrect inferences about the causal relations between the observed variables. For example, imagine measuring the variables X and Y and detecting what appears to be a causal relation from X to Y, for example because every event x is followed by an event y with some time lag, but not the other way around. This is insufficient information to infer a direct causal connection from X to Y, since there might be a third causally relevant variable, Z. Assuming Z remains unobserved, the purely statistical information about the relationship between X and Y is insufficient for singling out which of the three causal models shown in fig. 7 is correct. Together, they form the *Markov equivalence class* of this model. A Markov equivalence class is a set of DAGs which share the same set of conditional independencies between their variables. To identify the correct causal model in such cases, numerous methods have been found, and which one is applicable depends on the type of dataset and which assumptions can be made about it. For example, the *average treatment effect* can be used in cases where interventional data is available. This will be discussed further in section 4.2.

Causation at Multiple Levels For the current purposes, the aim is to look at the potential for studying causal effects at different levels. This gives a different causal model at a different level of detail, but both are equally true. In the same way as one might increase the level of detail through reductionism, it is also possible to go the other way, and to treat a composition of causal units and effects at one level as higher-order causal units at a different level. A challenge, then, is how to quantify this causal effect at higher levels.

In the case of simplicial structures, one of the underlying assumptions for this project was that the simplices themselves could serve as a type of causal unit. However, the concepts from causal inference largely assume that the causal units are well separated and both causal events and outcomes are clearly defined. This is not the case in the simplicial structures in biological neural networks, where different simplices are often largely overlapping, and both external input and outputs are present at many places between the source and the sink neurons. Further, whereas a spike event in a single neuron is a clearly defined event, what constitutes "activity" in a larger simplex is not clear. For these reasons, it is far from trivial exactly how one should apply the methods from causal inference when analysing the causal effects in simplicial structures.

The approach here chosen is thus not necessarily the only possible way of quantifying the causal effects in these structures. When looking at a simplex as a causal unit, the emphasis here has been on the informational flow from the source to the sink, which were treated as

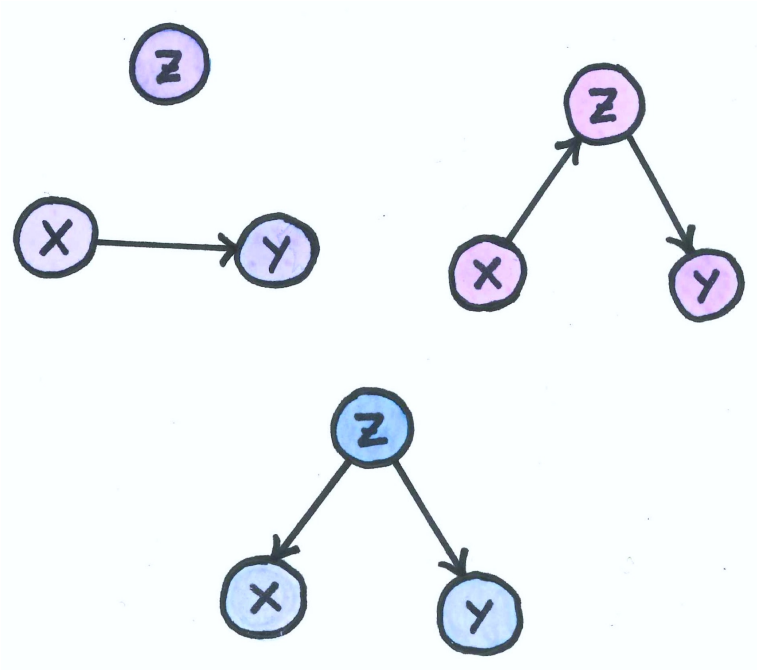


Figure 7: Three causal diagrams in the same Markov equivalence class. With only observational data with the final outcome of the three variables, it is impossible to determine which is the correct causal model.

the two informational "bottlenecks" in the structures. Thus, by investigating the observable effects of the activity in the source neurons on the activity in the sink neurons in different simplices, this should work as a proxy for the causal effects between simplices, as well as for the causal significance of the different changeable variables of simplices.

2.4 Machine Learning - Neural Networks

The main machine learning method used in this project is called *graph neural networks* (GNNs). This is often described as a generalisation of convolutional neural networks applicable to graphs. We shall not go too deeply into the details regarding CNNs, but shall limit ourselves so an overview of regular deep neural networks and the learning algorithms pertaining to these, which also form the backbone of graph neural networks. A brief overview of the main learning algorithms of GNNs will then be given.

2.4.1 Feed Forward Neural Networks

For simplicity, we shall consider a simple *feed forward neural network* with only a single input data-point consisting of a vector $\mathbf{a^{(0)}} = [a_1^{(0)}, a_2^{(0)}, ..., a_n^{(0)}, ..., a_N^{(0)}]$. The ground truth which we wish to predict is also a vector $\mathbf{t} = [t_1, t_2, ..., t_f, ..., t_F]$. The aim is thus to train a network to perform the mapping $\mathbf{a^{(0)}} \to \mathbf{t}$. This is done by passing the input vector through multiple layers of weights and biases, with activation functions after each layer. Each layer can be thought of as consisting of a set of K nodes. The output from layer l is denoted $\mathbf{a^{(l)}} = [a_1^{(l)}, a_2^{(l)}, ..., a_k^{(l)} ..., a_K^{(l)}]$. Each node in a given layer is connected to all the other nodes

in both the precesing and the succeeding layer, but there are no connections between the nodes in the same layer. The layers between the input and output are referred to as *hidden layers*, and the dimensionality of these are optional and independent of the dimensionality of the input and the output. For a network consisting of L layers, *L-2* of these are thus hidden. The weight matrix at layer l is denoted $\mathbf{W^{(l)}}$, the elements of which are $w_{jk}^{(l)}$. Each layer l also has a bias vector $\mathbf{b^{(l)}} = [b_1^{(l)}, b_2^{(l)}, ..., b_k^{(l)} ..., b_K^{(l)}]$. The output vector from a given layer before the activation function σ is applied is denoted $\mathbf{z^{(l)}}$. This is of the same dimensionality as $\mathbf{a^{(l)}}$, and is given by

$$\mathbf{z^l} = \mathbf{a^{l-1}W^l} + \mathbf{b^l}$$

The value of the node k in layer l is given by:

$$a_k^{(l)} = \sigma(z_k^{(l)}) = \sigma\left(\sum_{j=1}^{J} a_j^{l-1} w_{jk}^l + b_k^l\right)$$

Here, j refers to the index of the nodes in the previous layer, the total of which is J.

The complete weight matrix at level l is given by

$$\mathbf{W^{(l)}} = \begin{bmatrix} w_{11}^{(l)} & w_{12}^{(l)} & \dots & w_{1k}^{(l)} & \dots & w_{1K}^{(l)} \\ w_{21}^{(l)} & w_{22}^{(l)} & \dots & w_{2k}^{(l)} & \dots & w_{2K}^{(l)} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ w_{j1}^{(l)} & w_{j2}^{(l)} & \dots & w_{jk}^{(l)} & \dots & w_{JK}^{(l)} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ w_{J1}^{(l)} & w_{J2}^{(l)} & \dots & w_{Jk}^{(l)} & \dots & w_{JK}^{(l)} \end{bmatrix}$$

A visualisation of this mapping is shown in fig. 8.

Neural networks are trained to minimise some *cost function*. This determines the metric used to evaluate the quality of the model, and is usually a measure of the error. Training the network is thus equivalent to updating the model in order to minimise this cost function. Different cost functions are applicable for different purposes, but the most common for continuous values, and the one used here, is the *mean squared error* (MSE). With the final prediction being $\mathbf{a^L} = [a_1^L, a_2^L, ..., a_f^L, ..., a_F^L]$ and the ground truth $\mathbf{t} = [t_1, t_2, ..., t_f, ..., t_F]$, the final MSE loss is given by:

$$C = \frac{1}{F}\sum_{f=1}^{F}(t_f - a_f^L)^2.$$

There is a great range of different activation functions which can be used. For the model used in this project, ReLU was used, and this is given by

$$\sigma(z) = \begin{cases} z \text{ for } z > 0 \\ 0 \text{ for } z \leq 0, \end{cases}$$

with range $[0, \infty)$ and its derivative by

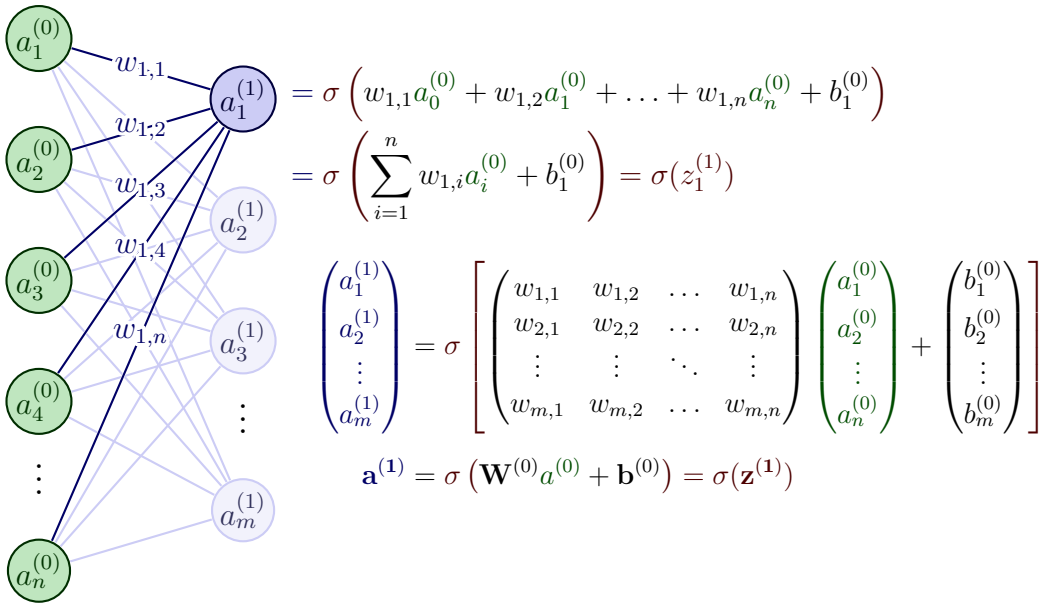


Figure 8: A feed forward neural network showing the input layer and the first hidden layer along with the corresponding calculations. Figure adapted from [42].

$$\sigma'(z) = \begin{cases} 1 \text{ for } z > 0 \\ 0 \text{ for } z \leq 0. \end{cases}$$

Note that it is common to not use an activation function at the final layer, since the ranges of these do not always correspond with the desired output range.

When training the model, for each iteration, each weight (and bias) is updated by taking the derivative of the final error with respect to this weight (and bias), such that the new weights and biases are given by

$$w_{jk}^{(l)} \Leftarrow w_{jk}^{(l)} - \eta \frac{\partial C}{\partial w_{jk}^{(l)}},$$
$$b_{k}^{(l)} \Leftarrow b_{k}^{(l)} - \eta \frac{\partial C}{\partial b_{k}^{(l)}},$$

where η is a small number known as the *learning rate*. In practice, this is done through the *backpropagation* algorithm, which is used to update both the weights and the biases. First, the error in the final layer is calculated, and from there, the weights and biases are updated in reverse order according to the following algorithm.

From the chain rule, the error in node k in the output layer is given by:

$$\delta_k^L = \frac{\partial C}{\partial a_k^L}\frac{\partial a_k^L}{\partial z_k^L} = \frac{\partial C}{\partial a_k^L}\sigma'(z_k^L).$$

In matrix notation, we have

$$\delta^{\mathbf{L}} = \nabla_{a_L} C \odot \sigma'(\mathbf{z^L})$$

The errors in each layer are then calculated iteratively in reverse order as a function of the errors in the succeeding layer according to the following expression. (The full derivation of which can be found in a range of different sources on the topic, such as [43].)

$$\delta^{\mathbf{l}} = \delta^{\mathbf{l+1}}(\mathbf{W}^{\mathbf{l+1}})^{\mathbf{T}} \odot \sigma'(\mathbf{z}^{\mathbf{l}})$$

Here, $\odot$ denotes the Hadamard product. The gradients of the weights and biases are then given by

$$\frac{\partial C}{\partial w_{jk}^{(l)}} = a_k^{(l-1)}\delta_j^l$$

$$\frac{\partial C}{\partial b_k^{(l)}} = \delta_k^l$$

In matrix notation, the updates are given by

$$\nabla \mathbf{W}^{\mathbf{l}} = (\mathbf{a}^{\mathbf{l-1}})^{\mathbf{T}}\delta^{\mathbf{l}},$$

$$\nabla \mathbf{b}^{\mathbf{L}} = \delta^{\mathbf{l}},$$

And the updated weights and biases are thus

$$\mathbf{W}^{\mathbf{l}} \Leftarrow \mathbf{W}^{\mathbf{l}} - \eta \nabla \mathbf{W}^{\mathbf{l}},$$

$$\mathbf{b}^{\mathbf{l}} \Leftarrow \mathbf{b}^{\mathbf{l}} - \eta \nabla \mathbf{b}^{\mathbf{l}},$$

In practice, this is done using *autograd* (backward automatic differentiation) i `Pytorch`. Finally, a commonly encountered problem in deep learning is that the gradients of the weights either become extremely large (exploding) or diminishingly small. It has been found that initialising the weights in certain ways can sometimes help in mitigating these problems, as well as improving the convergence rate of the solution. The intuitive reason for this is that the input to each node is a weighted sum of all of the outputs from the preceding layer, so somehow scaling the weights with the number of nodes in the layers makes intuitive sense.

In the model here used, the weights were all initialised using what is known as the Kaiming-He initialisation method, as described in [24]. The weights in layer l are initialised according to the following normal distribution

$$W^l \sim \mathcal{N}\big(0, \sigma^2\big).$$

Where the standard deviation is given by

$$\sigma = \sqrt{\frac{2}{J}},$$

where J is the number of nodes in the preceding layer.

2.4.2 Graph Neural Networks

An important limitation of standard feed forward neural networks is that the ordering of the input matters, such that reshuffling the input $\mathbf{X}$ will in most cases change the output. Depending on the intended application, this may or may not be a problem. A typical example of where this *is* a problem is with images, where we often want the prediction, say the classification of the object in the image, to be both rotationally and translationally invariant, such that the prediction remains the same regardless of the object's position and orientation in the image. To deal with this problem, convolutional neural networks (CNNs) were introduced. Another data-type which requires a similar form of invariance is graphs, where the same information can be represented in a multitude of ways. To deal with this, graph neural networks (GNNs) have been introduced, and these can be seen as a generalisation of CNNs applicable to graph data.

As above, we define a graph as a set of nodes and edges, where both the nodes and the edges can have a set of corresponding attributes, both of which can be represented in vector or matrix form. For a graph of N nodes, each with F features, the nodes are represented by the matrix $\mathbf{X} \in \mathbb{R}^{N \times F}$. Similarly, for edges with S attributes, for each edge from node i to j, there exists an edge vector $\mathbf{e}_{ij} \in \mathbb{R}^S$.

The interactions between the nodes are dependent on the edges between them, and these can be represented in the binary *adjacency matrix* $\mathbf{A}$, where for each edge $\mathbf{e}_{ij}$, there is a corresponding non-zero element in $\mathbf{A}$. There is a great variety of GNN structures, and which one to use depends on the task at hand. The current exposition will be limited to the task relevant for the current project, which is to go from one set of node embeddings $\mathbf{X} \in \mathbb{R}^{N \times T}$ to another, $\mathbf{Y} \in \mathbb{R}^{N \times N}$, where both the adjacency matrix and the edge features are unknown.

For this purpose, the *message passing algorithm* of GNNs will be useful. This consists of three steps: sending the message, aggregating the messages, and updating the node embeddings. The general formula for a message passing GNN is given by

$$\hat{\mathbf{x}}_i = \gamma^{(k)} \left(\mathbf{x}_i, \square_{j \in \mathcal{N}(i)} \phi \left(\mathbf{x}_i, \mathbf{x}_j, \mathbf{e}_{ji} \right) \right) \tag{2}$$

Here, $\mathcal{N}(i)$ represents all the neighbours of node i. In the case of a directed graph, they are the nodes from which there is an incoming edge to node i. At each step, the node i thus receives a message from every other node j from which it has an incoming edge. This message depends on the node attributes x_i and x_j, and possibly the edge attributes of e_{ji}. Each message is passed through a neural network, often referred to as an MLP (multi-layer perceptron), and here denoted by ϕ. All incoming messages are then aggregated according to some aggregation function, here denoted $\square_{j \in \mathcal{N}(i)}$. The exact type of aggregation will depend on the purpose, but common choises include sum, average or concatenation. Finally, the update function γ, which is also an MLP, takes the previous embedding $\mathbf{x}_i$ along with the output from the aggregation function, to learn a new node embedding $\hat{\mathbf{x}}_i$. All of this is easily implemented using the `MessagePassing` class from `PytorchGeometric`. At the end of each epoch, the predicted embeddings $\hat{\mathbf{x}}_i$ are then compared with the ground truth $\mathbf{y}_i$, and the loss calculated according to the cost function chosen. The learnable parameters are the weights and biases in the two MLPs ϕ and γ, and these are updated using the same backpropagation algorithm as for regular feed forward neural networks. By iteratively

performing this message passing step on a range of different graphs, the model learns to perform the mapping $\mathbf{X} \in \mathbb{R}^{N \times T} \rightarrow \mathbf{Y} \in \mathbb{R}^{N \times N}$.

2.5 Dataset

The datasets analysed in this project consisted of simulated activity in artificial neural networks. These datasets were generated using a computational implementation of a generalised linear model, and the connectivity graphs were chosen according to the purposes. In the current section, the model and the simulator will be described in detail, followed by a section covering the different connectivity graphs studied. Finally, some of the shortcomings and limitations of the dataset will be discussed.

2.5.1 Modelling Neural Dynamics

Given the many challenges related to the different types of recordings in experimental neuroscience, being able to computationally model neural dynamics in systems where all variables of interest can then either be tuned or easily obtained is of immense value. There are numerous approaches to modelling neuronal dynamics, and which one to use depends on the purpose in question. In our case, the detailed descriptions of the ion fluxes, the change in the membrane potential or the exact shape of the action potential is unimportant. Instead, it is sufficient to treat an action potential as an event with a binary value. In such cases, one of the simplest models one can use is the *(leaky) integrate-and-fire model* of neural dynamics. However, this model is not optimal when the aim is to accurately model the spike statistics of the neurons, so a class of more generalised models have been introduced, called *generalised linear models* (GLM). A brief overview of both types of models are given below.

Leaky Integrate and Fire Models Integrate and fire models are in essence very simplified descriptions of the dynamics leading up to an action potential being fired in a single neuron. It takes the shape of a differential equation describing how incoming signals are integrated and transformed into a changing membrane potential, and the subsequent firing of an action potential and resulting discharge. In its simplest form, it describes the evolution of the membrane potential and the threshold for the firing of an action potential. The solution to this equation is given by

$$u(t) - u_{\text{rest}} = \Delta u \exp\left(-\frac{t - t_0}{\tau_m}\right) \qquad \text{for } t > t_0$$

where the membrane potential at time $t = 0$ is

$$u(t_0) = u_{\text{rest}} + \Delta u$$

such that Δu is some constant defined thereafter. τ_m is the characteristic time of the decay of the membrane potential, which is typically in the range of 10 ms [20].

The leaky integrate and fire model is a highly simplified model of actual firing dynamics, and neglects many salient features of neuronal dynamics. Two crucial omissions are the dependence on the state of the receiving neuron and the spiking history. The model assumes

that the inputs are integrated linearly, independently of the actual state of the post-synaptic neuron. Additionally, the membrane potential is simply reset to the resting potential after each firing event, so no memory of previous spikes is retained. These, along with a range of other omissions (see [20], section 1.4 for a more extensive list and discussion of these), may or may not be a problem, depending on the intended use of the model. It is possible to include additional variables to account for the dependency of the activity on the spike-history of the neuron, thereby allowing for adaptability and refractoriness, as well as non-linearities in the synaptic responses. For even more complex descriptions of the neural dynamics, so-called generalised leaky integrate-and-fire (GLIF) models can be used. These can include a range of other variables describing the adaptability of the spike-responses and other dynamic dependencies within the networks.

Generalised Linear Models Generalised linear models (GLMs) can be seen as a further extension of these. These are useful when the main emphasis is on the spike statistics of the neurons, which they capture more accurately by including a stochastic element in the spike responses. In essence, GLMs consist of a linear filter (which is based on the solution to a GLIF model), a link function and a probability distribution [46]. The input signal to each cell is described by the set of linear filters, including a stimulus filter, a post-spike filter capturing the dependencies of the activity on the spike-train history of the neuron itself and a coupling filter which gives the dependencies on the recent spike-histories of its pre-synaptic neurons. In a Bernoulli GLM, which is what was used in this project, the input is then passed through the link function, which is a non-linear function converting the membrane potential resulting from the input activity to an instantaneous spike probability. Finally, a probability distribution is applied to this probability, determining whether or not a spike event occurs in the neuron in question [44].

The main difference between such models and the generalised integrate-and-fire models is the stochastic element arising from the probability distributions. Further, the model contains a range of tunable parameters which can be fit to data, such that GLM models can be used to predict future activity in individual neurons or larger networks given previous activity and new stimuli. Undoubtedly, these models are still gross simplifications of biological neural networks in their full complexity. Perhaps most importantly, they completely ignore all influences coming from other factors than the chemical synapses, such as the neuromodulators and neuropeptides in the extracellular environment and effects from other neurons connected through electrical synapses.

2.5.2 Neural Activity Simulation

The activity in the networks here studied was simulated using a generalised linear model, specifically, the Bernouilli GLM as presented in [32] and implemented using an early version of the `spikeometric` package in Python, though with some adjustments. The code used to generate the data is available at [28].

The activity in the network is represented as a matrix $\boldsymbol{X}$ containing the spike-trains of each of the neurons in the network. This is encoded as zeros and ones, where each element represents whether the neuron spiked at a given time-step or not. Thus, for a network of N neurons with activity generated over a period of T time-steps, $\boldsymbol{X}$ is a matrix of dimension

$N \times T$. The network itself is represented as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of N nodes, and $\mathcal{E}$ the directed edges between them in the network.

These connections are also represented in the weighted connectivity matrix $\mathbf{W_0}$, whose elements represent the weighted synaptic connections between the neurons in the network. The elements in this matrix are thus given by the elements of $\mathcal{E}$, as well as a weighting function $\phi : \mathcal{E} \to \mathbb{R}$. Thus,

$$(W_0)_{i,j} = \begin{cases} \phi(i,j) & (i,j) \in \mathcal{E} \\ 0 & \text{otherwise} \end{cases}$$

Note that this connectivity matrix can either be interpreted as the structural or the effective connectome. If taken as the structural connectome, each element represents a single synapse, meaning that we only allow a single synapse to exist in each direction for any pair of neurons. If taken as the effective connectome, we may allow for multiple synapses to exist in the same direction for the same pair of neurons, and the elements of $\mathbf{W_0}$ would represent the sum of the effect mediated through all the synapses between the pair. For simplicity, we here adopt the former interpretation. Note that this means that since other factors, such as the neuromodulatory environment or the electrical synapses, are not taken into account, and $\mathbf{W_0}$ is kept constant throughout the entire simulation, $\mathbf{W_0}$ represents both the structural and the effective connectome in this case.

Further, the influence from an individual spike may last for more than one time step. The influence is set to be a decaying function over the *coupling window* of c_w time steps. This is implemented by defining a connectivity filter tensor with size $N \times N \times c_w$, such that the influence of neuron i on neuron j at time-step t after a spike-event at neuron i is given by

$$W_{i,j}(t) = \begin{cases} (W_0)_{i,j}\, e^{-\beta t \Delta t} & \text{if } t < c_w \\ 0 & \text{if } c_w \leq t \end{cases} \tag{3}$$

where β is the (tuneable) decay rate of the weights and Δt is the time-step size used. This was kept constant at 1 ms. The exponential term corresponds to the coupling function $c(t)$, determining the rate of the decay.

The dynamics of the synaptic interactions are here represented by a current-based model. The interaction between the neurons is simply represented as a current injected into the postsynaptic neuron which directly affects the membrane potential, without regard for the changes in the conductance of the neuron's membrane. To simplify further, this change in the post-synaptic current is modelled as a step-function, rather than as a time-varying function, thereby assuming that the synaptic strength is constant throughout the interaction.

In addition to the incoming activity at each time step, a neuron's chance of firing is also dependent on its absolute and relative refractory periods, and thus on its previous activity. As mentioned in section 2.1.1, the *absolute* refractory period is the period after a spike during which it is impossible for a neuron to spike again, whereas the *relative* refractory period is the succeeding period during which a higher input activation is needed to excite a new action potential. This refractory filter is implemented as a function of the time t since the last spiking event, and is given by

$$r(t) = \begin{cases} a & \text{if } t < A_{ref} \\ re^{-\alpha t \Delta t} & \text{if } A_{ref} \leq t < A_{ref} + R_{ref} \\ 0 & \text{if } A_{ref} + R_{ref} \leq t \end{cases}$$

To summarise, the following parameters are included in the model, listed along with their default values:

- $\theta = 4.3$ - the activation threshold above which the neuron spikes with a probability above 50%
- $\Delta t = 1$ ms - the time-step size, in ms
- $c_w = 10$ ms - the length of the coupling window
- $\alpha = 0.5$ - the decay rate of the negative activation during the relative refractory period
- $A_{ref} =$ -100 ms - the absolute refractory period of the neurons
- $a =$ -3 - the negative activation added to the neurons during the absolute refractory period
- $R_{ref} =$ -30 ms - the relative refractory period of the neurons
- $r =$ -7 - the negative activation added to the neurons during the relative refractory period
- $\beta = 0.2$ - the decay rates of the weights

The activity in the network is generated at each time-step, and can be split into three steps. First the input each neuron receives from all incoming edges, as well as random input noise, it calculated. This is implemented using the `MessagePassing` class from `torch_geometric`. Second, this activity is passed through a nonlinear sigmoid function σ to get the probabilities of firing. Finally, whether each neuron fires or not is determined by sampling from the Bernoulli distribution of the probabilities. These three steps are summarised as:

$$g_i(t+1) = \sum_{\tau=0}^{T-1} \left(X_i(t-\tau) r(\tau) + \sum_{j \in \mathcal{N}(i)} (W_0)_{j,i} X_j(t-\tau) c(\tau) \right) + \mathcal{E}_i(t+1)$$

$$p_i(t+1) = \sigma\left(g_i(t+1) - \theta\right) \Delta t$$

$$X_i(t+1) \sim \text{Bernoulli}\left(p_i(t+1)\right)$$

The sigmoid function is given by

$$\sigma(x) = \frac{1}{1+e^{-x}}.$$

The Bernoulli distribution is a binary and discrete probability distribution, used to model random experiments where the outcomes are 0 or 1. The distribution is characterised by the variable p, which is the probability of the random variable taking the value 1, or, conversely, by the probability of it taking the value 0, denoted $q = 1 - p$. The probability of the variable X taking the value x, for $x \in \{0, 1\}$ is thus given by

$$P(X = x) = p^x(1 - p)^{(1-x)}.$$

This distribution has a mean of p and a variance of $pq = p(1 - p)$.

2.5.3 Connectivity Graphs

The connectivity graphs used in these investigations were all computationally generated, and their resemblances to biological neural networks are thus limited. Three main types of binary graph structures were investigated, for different purposes, but also for comparison between them. For each combination of variables, 200 datasets were generated. The distribution of inhibitory and excitatory neurons was set to 50/50. This was to ensure stability in the activity, and given that the networks studied were very small, setting the ratio to 20/80 with different strengths made the activity overly dependent on the distribution of the weights. In many cases, this led the activity to either explode or die out, so the 50/50 ratio was used instead. The synaptic weights were generated in the same way for all three types, with different randomly chosen weights for each dataset. The absolute values of the edge weights were all taken from a random uniform distribution $\mathcal{U}(0, \frac{6}{\sqrt{N}})$, where N is the number of neurons in the network. This scaling was found to be necessary to ensure stable activity in the networks regardless of their size.

Table 1: The number of neurons and edges in simplices of different dimensions

Dimension	Neurons	Edges
2	3	3
3	4	6
4	5	10
5	6	15
6	7	21
7	8	28
8	9	36
9	10	45

Idealised Simplices The simplest networks studied consisted of only one directed simplex of a given size, of which all the neurons in the network formed a part. The connectivity matrix is thus an upper triangular matrix with the diagonal as 0 (no self-loops). These very simple

structures are of course extremely unrealistic as components of real biological networks, but this simplicity greatly facilitates the functional analysis for which they were used. Contrary to the directed simplices found in more integrated networks, these isolated simplices have no confounding variables between the source and and sink, and this makes it possible to identify the functional potential of the simplex in its purest form. These graphs can thus be seen as idealised models used to isolate the functional effect the activity in the source neuron has on the activity in the sink neuron. When instantiated in other networks, the actual effect will of course be influenced by a range of other factors, and will differ from network to network. The number of edges in each network is determined by the number of nodes, and are listed in table 1.

Small World The graph structure used as the model for biologically realistic networks were small-world networks, also known as *Watts–Strogatz* networks. Small-worldness involves that the nodes are relatively sparsely connected, but the average geodesic (shortest path length) between any two nodes is nevertheless relatively short. It has been found that biological neural networks to a great extent share these properties. The ubiquity of a certain structure in biological systems might indicate that it has been adapted for evolutionarily, and further that this is because it serves certain beneficiary functions. In the case of small-world networks, it has been hypothesised that they facilitate effective information flow and thereby reduce the energy expenditure [51]. It has also been hypothesised that small-world networks are more robust, since small perturbations to the networks, such as adding or removing individual nodes or edges, rarely cause significant changes to the global properties of the networks [52].

The graphs were generated in python using the `watts_strogatz_graph` method from `networkX`, which takes three parameters, N, k and p. Given these numbers, the algorithm for generating the graph is as follows. First, a ring containing N nodes is created. Next, each node in the ring is connected to its k nearest neighbours (or (k-1) if k is odd.) Finally, for each edge (u, v) in the network, this is replaced with probability p with another edge (u, w), where w is a randomly chosen node in the network with $w \neq v$. This has the effect of creating shortcuts in the network. Following from this algorithm, the number of edges will be the same for each graph containing the same number of nodes. The number of nodes, edges and k-values for the different graphs studied are listed in table 2. For reasons of scalability, the k-values used were dependent on the number of neurons, N, and given by:

$$k = \sqrt{N}$$

rounded off to the nearest integer.

It should be noted that although such small-world graphs share certain properties with biological neural networks, they are far from realistic reproductions. The main commonalities are the short average path length and the high degree of clustering. The main limitation is the degree distribution. The degree of a node is defined as the number of other nodes to which it is connected, so the degree distribution is the probability distribution of the degrees of all the nodes in the network. For small-world networks, this distribution is centred with a sharp peak around k, meaning that the connectivity of the network is fairly homogeneous, with most of the nodes having similar degrees. Such patterns are rarely seen in real networks,

Table 2: The number of neurons, edges and the k-values in small-world networks of different sizes

Neurons	k	Edges
10	3	20
15	4	60
20	4	80
25	5	100
30	5	120
40	6	240
50	7	300
60	8	480
70	8	570

so is a substantial limitation to how realistic these networks are. In contrast, many real networks are scale-free, meaning that the degree distribution follows a power law. This is caused by new nodes having a *preferential attachment* to nodes of already high degree. The nodes with high degree thereby end up serving as hub nodes between clusters of less densely connected nodes. However, a limitation of these is that they fail to reproduce the high degree of clustering seen in many real networks, which the small-world networks do indeed reproduce, so neither of the graphs can be treated as fully realistic as reproductions of biological neural networks [51].

Erdős–Rényi Erdős–Rényi graphs are a type of a random graphs, and these were generated using the corresponding function from `networkX`. This function takes two arguments, N and p, where N is the number of neurons and p the probability that any possible edge in the graph is present. This introduces an element of stochasticity in the actual number of edges in the graph, where the average number of edges in a graph with N nodes is given by $(N \times N - N) \times p$. For comparability, the values used for p were chosen to match the number of connections with those in the small-world networks containing the same number of nodes. Thus, the value of p was adjusted for graphs of different sizes according to the following relation.

$$p = \frac{E}{N \times N - N}$$

Where E is the number of edges in the small-world graph with the same number of nodes. The resulting values of p, as well as the actual average number of edges in each dataset are listed in table 3.

Table 3: The number of neurons, the probabilities p and the average number of edges in each dataset for the Erdős–Rényi networks

Neurons	p	Avg. edges
10	0.22222	20.21
15	0.28571	60.54
20	0.21053	80.78
25	0.16667	100.02
30	0.13793	119.50
40	0.15385	240.04
50	0.12245	297.39
60	0.13559	477.93
70	0.11801	567.91

Metrics For each graph, both their complete and maximal simplicial complex were calculated and represented in Hasse diagrams. This gives a complete description of the entire graph. The "simplicial properties" of the individual nodes were also computed. This included a count of how many simplices of each dimension the node served as a source, mediator and sink. In addition to these simplicial properties, the in- and out-degrees of the nodes were also considered. The in-degree of a node is simply the number of incoming edges to that node, and the out-degree the number of outgoing edges. The degree of a node is the sum of these two. These properties are of course correlated with the simplicial role of a node, but it is not a symmetric relation. Being a source in an n-simplex automatically implies an out-degree of at least n, but a node can have the same out-degree without being the source of an n-simplex. Since the aim is to study the functional role of the *simplicial structures*, and not just that of the clustering and degree, one must control for the effect of the degree of the nodes.

2.5.4 Limitations

The datasets studied are to some extent realistic, but some realisticness is necessarily sacrificed for the sake of simplicity and interpretability. This is an inevitable part of any model in science, but it is important to be conscious of their shortcomings. Regarding the connectivity graphs, the 50/50 ratio of excitatory and inhibitory neurons and their equal distributions in the absolute value of their synaptic strengths are known falsehoods. Furthermore, the distribution of excitatory and inhibitory neurons is not random in real neural networks, although it is assumed to be so here. For example, neurons which send signals to other regions of the brain tend to be excitatory, whereas inhibitory neurons tend to have shorter range and often form part of circuits where their signals are looped back to earlier segments of the same circuit [62]. Regarding the activity simulator, this model only takes into account the activity which results from the synaptic input of chemical synapses, so includes no consideration for

changes in the extracellular environment nor effects from electric synapses. Although these simplifications are not believed to be of major significance, it is something worth keeping in mind, especially when considering the functional analysis performed on the simulated data.

2.6 Statistical Toolkit

When analysing the results obtained, a range of different methods from statistics were used. An overview of these are given here.

2.6.1 Standard Metrics

Assume a dataset consisting of N datapoints $x = x_1, x_2, ..., x_N$. The arithmetic mean of the dataset is given by

$$\bar{x} \equiv \frac{1}{N} \sum_{i=1}^{N} x_i.$$

Another useful metric is the *variance* in the data, which is defined as the average distance squared between the datapoints and the mean. This is given by:

$$\begin{aligned} V &\equiv \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2 \\ &= \overline{x^2} - \bar{x}^2 \end{aligned}$$

From this, we can define the *standard deviation*, which is

$$\sigma \equiv \sqrt{V}$$

This is measured in the units of the data itself, and is commonly used to evaluate how unusual a datapoint is. Assuming the data is normally distributed, 68.3 % of the datapoints will lie within the range of one standard deviation around the mean, 95.4 % of the datapoints within the range of two standard deviations around the mean and 99.7 % within the range of three standard deviations around the mean. Whereas the standard deviation of a dataset represents the average distance of each datapoint from the mean of the dataset, so describes the spread of individual datapoints within a dataset, the *standard error of the mean*, or simply the *standard error*, is a measure of the variability of the sample mean from the true population mean, so tells us something about how accurate the sample mean is relative to the population mean. This is given by

$$SE \equiv \frac{\sigma}{\sqrt{N}}.$$

Further, it is common to use a slightly adjusted form of the variance and standard deviation, given by

$$
\begin{aligned}
V_{N-1} &\equiv \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})^2 \\
&= V_N\frac{N}{N-1} \\
\sigma_{N-1} &= \sqrt{V_{N-1}} \\
&= \sigma_N\sqrt{\frac{N}{N-1}}
\end{aligned}
$$

This is motivated by the fact that our datasets are always subsets of the complete (and ideal) dataset obtained for $N \to \infty$, and the variance and standard deviations which we calculate are really descriptions of the subsets studied, whereas what we actually want is the variance and standard deviation of the complete (and infinitely big) dataset. The *N-1* versions of these equations are better estimates of the true variance and standard deviations, which is what we wish to approximate.

Further, it is often useful to study the relationships between different datasets. Assuming another dataset with the same number of datapoints as the original, $y = y_1, y_2, ..., y_N$, the covariace between x and y is defined as:

$$
\begin{aligned}
\mathrm{cov}(x,y) &\equiv \frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y}) \\
&= \overline{xy} - \bar{x}\bar{y}
\end{aligned}
$$

This tells us something about how the parameters are related to each other, but has the disadvantage that it also depends on their respective ranges, and has units. A more informative metric is thus the *correlation coefficient*. This is in the range [-1, 1], where -1 indicates that the datasets are perfectly anti-correlated, 0 that the datasets are uncorrelated, and 1 that they are perfectly correlated. The correlation coefficient between x and y is defined as

$$
\rho_{xy} \equiv \frac{\mathrm{cov}(x,y)}{\sigma_x\sigma_y}.
$$

When presenting the obtained results, it is common to give the calculated value along with its *confidence interval*. The confidence interval is the error bars on the results, and it is common to set this to being within one, two or three standard deviations. The higher the error, the more certain one can be that the true value lies within the range stated (see above for the certainties corresponding to the different ranges), at the cost of precision.

2.6.2 The Normal Distribution

In statistics, how the variables are distributed is an important factor for determining the appropriate statistical tools to be used. A statistical distribution shows the number of observations for each value of the variable in question, or for each group of variable values. It is thus a useful tool when studying either the spread of the sample values or the probability of observing a certain value. Commonly used metrics for characterising distributions are

their mean, variance, standard deviation and shape. For the current purposes, we shall limit ourselves to a brief introduction to the normal distribution.

The normal distribution, also called a *Gaussian* distribution, is a symmetric distribution with a single peak around the mean. The probability distribution of a normal distribution is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

where μ is the mean around which the distribution is centred, σ is the standard deviation and x is the independent variable. This distribution is commonly used to represent real-valued random variables, both in the natural and the social sciences. Part of its significance derives from the *central limit theorem*. This states that in many cases, given a variable X, if we take multiple subsamples of X and calculate their means $\hat{m}$, as the number of samples increases, the distribution of these means will itself approach a normal distribution.

2.6.3 Null Hypothesis Significance Testing

When testing hypotheses, a common method is the null hypothesis significance testing. The reasoning behind this method is based on the following rule:

Definition 2.2: Rare Event Rule for Inferential Statistics

"If, under a given assumption, the probability of a particular observed event is extremely small, we conclude that the assumption is probably not correct." [57, p. 393]

In essence, this means that we analyse the data in order to be able to distinguish between events which could plausibly occur by chance and those that are highly unlikely to occur by chance. The method proceeds as follows. Assume we have an initial hypothesis H, which states that some variables are correlated. We then create a *null hypothesis*, H_0, which is the negation of this, namely that the variables are *not* correlated. This null hypothesis must be precisely defined in terms of a numerical value. In this case, it would state that the correlation coefficient $c = 0$. Next, we formulate an *alternative hypothesis*, H_1, which must imply that H_0 is false, although it need not be its perfect negation. This is because the alternative hypothesis can be either one-sided or two sided. If two-sided, it would state that $c \neq 0$, but it could just as well state that $c < 0$ or $c > 0$. In our case, where the original claim which we wish to test is that there is a positive correlation, the natural choice would be $H_1 : c > 0$. However, note that H_1 need not be equal to H. We then assume that H_0 is true, perform the experiments, and analyse the data in question using some appropriate *test statistic*. We then calculate the probability of seeing the observed data given that H_0 is true, so the lower the value, the less likely it is that H_0 is true. The resulting number is commonly known as the *p-value*. Beforehand, we have decided on a *significance level* α, which is a probability threshold below which the null hypothesis H_0 will be rejected. Common choices are 0.01 and 0.05. The lower the value, the harder it is for the null-hypothesis to be rejected, but the more statistically significant is the potential rejection. The obtained *p-value* is then compared with α. If $p < \alpha$, the null-hypothesis is rejected. In the case of a one-sided H_1, some further investigation is then needed to test whether we are also justified in accepting

H_1, whereas when H_1 is two-sided, this follows directly from the rejection of H_0. However, if $p > \alpha$, we *fail to reject* the null hypothesis. It is of crucial importance to stress that this does *not* mean that we *accept* H_0. All we can say from such a result is that we have failed to show that H_1 is the case with the required degree of certainty [57].

Errors For any set of hypotheses and observations, there are four potential outcomes. We can correctly reject H_0, correctly fail to reject H_0, incorrectly reject H_0 or incorrectly fail to reject H_0.

Incorrectly rejecting H_0 is known as a type I error, and involves that we reject H_0 and accept H_1, when H_0 is in fact true. The rate of type I errors depends on the significance level chosen. For example, with a significance level of 0.05, in the worst case we will incorrectly reject H_0 5% of the time, so whenever $p < \alpha$, we can only be 95% certain that the detected effect is real and not just a statistical fluctuation.

Incorrectly failing to reject H_0 is known as a type II error, and means that we fail to reject H_0, and thereby implicitly reject H_1, when H_1 was indeed correct. The rate of type II errors is usually denoted by β, and is the result of numerous other things, such as the sample size, the standard deviation and the significance level. In general, there is a trade-off between type I and type II errors, where the rate of type II errors increase with decreasing α.

2.6.4 Test Statistics

The general form of the test statistic is to calculate some test score on the following format

$$\frac{\text{Observed data} - \text{Expected data if } H_0 \text{ is true}}{\text{Average variation in data}}$$

In cases where the true distribution is known to be a gaussian and the standard deviation σ of the entire population is known, it is common to use a *z-test*. When testing a single sample of n elements, where $\hat{x}$ is the sample mean and μ_0 the hypothesised population mean (i.e. the expected mean if H_0 is true), the *z-value* is given by

$$z = \frac{\hat{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}.$$

However, in many cases, the standard deviation is unknown, and the distribution might not be perfectly gaussian. Assuming n is sufficiently large, it is possible to instead use a *t-test*. This is given by

$$t = \frac{\hat{x} - \mu_0}{\frac{s}{\sqrt{n}}},$$

where s is the sample standard deviation. To find the *p-value*, one also need the number of *degrees of freedom, df*, which is defined as

$$df = n - 1.$$

For the current purposes, the test statistic which will be most applicable is the *analysis of variance*, commonly shortened as ANOVA [15]. This is a type of *F-test statistic* which is used to test to what extent a certain grouping of the samples helps to explain their values. The null hypothesis is that the means of the underlying distributions of all the groups with respect to the variable in question are identical, with the alternative hypothesis being that at least one of the means differ from the others. In other words, it is a test of whether the grouping is meaningful for the current explanatory purposes. It does this by comparing the variance in the observations of the different groups. It comes in a few varieties, though for our purposes only the one-way ANOVA will be needed. This allows us to compare the variance in a single independent variable within and between groups. The resulting *F-statistic* is given by

$$\begin{aligned} F &= \frac{\text{explained variance}}{\text{unexplained variance}} \\ &= \frac{\text{between-group variance}}{\text{within-group variance}} \end{aligned}$$

Let K denote the number of groups, n_i the number of observations in the i^{th} group, $\bar{x}_i$ the sample mean of group i and $\bar{x}$ the overall sample mean. The between-group variance is then given by:

$$\frac{1}{K-1}\sum_{i=1}^{K} n_i(\bar{x}_i - \bar{x})^2$$

Further, let x_{ij} denote the j^{th} observation in the i^{th} group and let N be the overall sample size. The within-group variance is then given by

$$\frac{1}{N-K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2.$$

In this case, the degrees of freedom under the null hypothesis are given by $d_1 = K-1$ and $d_2 = N-K$, hence the denominators. If F is higher than some critical value corresponding to the chosen significance level α, the null hypothesis is rejected. This critical level is a function of the degrees of freedom and α, and can be found in tables. Alternatively, statistics packages in python usually allow you to go directly to the *p-value* for a straightforward comparison with α. In cases where only two groups are considered, the t-test and the ANOVA are equivalent, with their relation being $F = t^2$. Note that rejecting the null-hypothesis in an F-test only tells us that *at least one* of the groups differ from the others, though it does not tell us which one. To determine this, further investigations with pair-wise comparisons are needed. This can for instance be used to see whether there is only a single group which stands out, or whether numerous groups are significantly different from the others. The ANOVA test was applied using the `f_oneway` method from `scipy.stats`.

Importantly, there are three conditions which need to be met for the ANOVA test to be valid [58]. Firstly, the observations must be independent, secondly, the observations should be approximately normally distributed within each group, and thirdly, the variances must

be fairly equal across the groups. The first condition is easily met in our case, but tests must be performed to check whether the other two conditions are met.

Without going into further details, whether the data is normally distributed can be tested using the Shapiro-Wilk [50] test for normality, which was done using the `shapiro` method from `scipy.stats`. The Levene [33] test can be used to check for equal variance between the groups. This was done using the `homoscedasticity` method from `pingouin`. In both cases, the null-hypothesis tested assumes that the samples are normally distributed or that the variance is equal, such that p-values *below* 0.05 means that the null hypotheses are rejected, and the assumptions are *not* met. For p-values above 0.05, the null-hypotheses are not rejected, and we assume that the data is normally distributed or that we have homogeneity of variances.

If the normality-criterion is not met, one can use the Kruskal-Wallis [30] test instead. This is a non-parametric version of the one-way ANOVA, and is similarly used to test whether there are statistically significant differences between the observations from different groups. Whereas the ANOVA tests whether the means of the groups are different, the Kruskal-Wallis looks at the median values. These are then ranked, and the test checks whether a random observation from one group is equally likely to be above or below random observations from the other groups. This was implemented using the `kruskal` method from `scipy.stats`. In cases where the data is normally distributed, but the groups have unequal variance, the one-way Welch ANOVA [61] can be used instead. This was implemented using the `welch_anova` method from `pingouin`. Finally, in cases where neither condition is met, one can use the non-parametric Mann-Whitney U test [23]. This test only works for pair-wise comparisons, and tests whether the underlying distributions of the two datasets are the same. This was implemented using the `mannwhitneyu` method from `scipy.stats`.

In retrospect, the Mann-Whitney U test was the method primarily used, so we describe this in more detail. This is another rank-based test, where the null-hypothesis being tested is that given two random samples, x and y, one from each distributions, the probability of x being ranked higher than y is equal to the probability of y being ranked higher than x. If this is rejected, one can infer that the distributions are different. This is done by calculating the so-called U-statistic, which can be done by focusing on either sample 1 or sample 2, giving two different, but equivalent, numbers, namely

$$U_1 = R_1 - \frac{n_1 (n_1 + 1)}{2},$$

$$U_2 = R_2 - \frac{n_2 (n_2 + 1)}{2}.$$

n_1/n_2 are the sample sizes of sample $1/2$, and R_1/R_2 the the sum of the ranks of sample $1/2$. When calculating the p-values, the smaller of the two values U_1 and U_2 is used. Note that the shapes of the two sample distributions are important for how the results should be interpreted. If the two distributions have the same shape, meaning that their variances are equal, a statistically significant result on the Mann-Whitney U test means that the *means* of the two distributions are different. In the cases where the shapes are different, a positive result means that their *mean ranks*, i.e. the average of the ranks of all the observations in

each sample, are different. However, both results implies a statistically significant difference between the distributions.

3 Research Purpose and Conceptual Challenges

3.1 Overarching Purpose and Preliminary Work

As discussed in previous sections, the challenges we are facing when trying to understand the dynamics of biological neural networks are substantial, and there is no agreement on what approaches are likely to be most conducive in this very broad research program. The current project falls under the category of *connectomics*, which, as discussed in section 2.1.2, is in itself a field with much uncertainty about how to interpret different types of connectivity and connecting this to different functions.

Further, instead of simply looking for correlations in the data, as is common in the field of neuroscience, our emphasis was on the causal dependencies in the networks. This is inspired by Pearl's argument about the deeper ontological status of such causal relationships, and the belief that they more closely "carve nature at its joints" than mere descriptions of correlations between the variables. We believe that establishing the underlying causal mechanisms of the systems studied will be a better starting point for true understanding and mechanistic explanations of the dynamics of the activity in the networks.

The initial purpose of this project was to investigate and potentially develop machine learning methods for causal discovery from spike trains in neural structures. In the lingo of connectomics, this falls under the category of *effective connectomics*. However, this aim was eventually abandoned in favour of a more investigative analysis of the function of said structures, the reasons for which will be elaborated on below.

The machine learning model used was developed by Sønstebø and Brunborg, and is available at [8]. This model is based on graph neural networks, and is capable of discovering the structural connectome from activity generated using the simulator discussed above. A challenge with this model is that it learns a very high-dimensional version of the connectome, where the connection between every pair of neurons is mapped and given equal importance. This might indeed be useful in some cases, where the aim is to study smaller network motifs, but it does not scale well for larger systems with greater number of neurons, both when it comes to the computational power needed and our ability to interpret the output. For larger systems, the scale makes it near impossible to assign function to individual neurons and synapses. The initial aim of the this project was to develop a model which, given the same input, would be capable of learning a lower-dimensional representation of the connectivity by identifying bigger structural units than individual neurons. As discussed above, the brain shows a high degree of clustering and modularity, where the assumption is that these structural parcellations are indicative of functional parcellations, since segregated areas with high degrees of internal connectivity will be receiving and outputting much of the same information with small variations. It would thus be useful if a tool could be found to determine this low-dimensional structure of clusters and their inter-cluster connectivity, rather than the complete connectome.

Since the aim was to find a low-dimensional representation of the effective connectome, it was also required that the units of choice could somehow be interpreted as *causal* units, and after lengthy considerations of potential structures, the choice fell on the directed simplices, partly because of their relative novelty and hence lack of previous research done on them. As discussed above, there have been some investigations into their potential significance

in neural networks, but as of now, relatively little has been established, so this is a field in need of more research. This lack of research and known functionality of the structures led us to shift the focus away from the development of a mere predictive tool, and move towards a more functional investigation of these structures. After all, there is much to be said against spending great amounts of time and resources on developing a machine learning model capable of predicting a structure which you have no idea how to use or interpret. Hence the primary purpose of the project ended up being to determine the functional significance of higher-order simplicial structures in biological neural networks.

3.2 Conceptual Analysis of Simplices

As mentioned, studying the structural connectivity of neural networks in terms of simplicial complexes is a very new approach. Consequently, hardly anything is known about what they might signify, nor of whether they should even be treated as structural units of any special significance.

There are two main arguments in favour of the pursuitworthiness of simplices, one theoretical and one experimental. The theoretical argument is that they intuitively seem like important information processing units. In isolation, they only have a single input source, so all the neurons in the simplex initially receive the same bits of information from the source neuron. Next, there is a step-wise information processing where the number of integration points (and thus the number of steps) depends on the dimensionality of the simplex. For each additional neuron, the potential complexity of the information processing increases. Finally, the sink neuron serves as another bottleneck through which the processed output must pass. In the same way as other interconnected clusters are believed to serve as functional units because of their shared input, it is natural to believe that simplices might play a similar role.

The experimental motivation, as presented in [47] and outlined in section 2.2 above, is that higher-order simplices are found in unexpected numbers in reconstructed neural networks, compared with random (Erdös Rényi) graphs with the same number of neurons and synapses. A problem with this motivation is that these results are insufficient to justify the claim that simplicial structures play a designated role and that these are the units which have been selected for. It is often the case that we observe features in biological organisms in unexpected numbers. However, to go from there to assuming that the features necessarily serve a function which has been selected for will be an example of an *adaptationist fallacy*, as discussed by Gould and Lewontin [22]. They argue that many features will have appeared as necessary byproducts of other features which have been selected for. As an analogy, they use the spandrels in St Mark's Basilica in Venice as an example of an architectural feature which might easily and incorrectly be interpreted as an intentional decorative feature, when in reality they are simply an unavoidable byproduct of the construction of the dome. This illustrates the importance of always taking into account other features which might be the true evolutionary advantage which has been selected for, and from which the feature under investigation might simply be a byproduct. In short, one should be very cautious with assuming function just from presence.

Regarding simplices, we know that high degrees of connectivity and clustering are ubiquitous in biological neural networks, and are fairly confident that these traits do indeed serve a purpose in themselves. It might well be the case that the simplices' disproportional

pervasiveness is simply a byproduct of the high degrees of connectivity in the networks.

Further, the directed simplices discussed in [47] are very strictly defined, even though one might expect that near-perfect simplices, with only a few edges missing, will be able to perform very similar computations. To make the argument that it is the perfect simplices which have been selected for and are therefore worthy of special treatment as functional units in the brain, further justification is needed. One approach would be to show that they are found in disproportionately large numbers even when the high degree of connectivity is corrected for, and that their numbers are more statistically significant than similar structures which one would also expect to follow as a byproduct from high levels of connectivity. Another result to strengthen the belief that directed simplices are somehow special would be to show that they are especially suited to perform certain computational functions which we believe to be common, or at least favourable, in the brain. A significant part of this project has been focused on the latter question, thereby also investigating the pursuitworthiness of higher-order simplices as a significant unit in biological neural networks. The ultimate aim would be to connect these structures to higher-level cognitive functions, but in the current project, we were looking at the neural dynamics at a much lower level. The best we could hope for was thus to find some lower-level functional properties which might plausibly be connected with higher-level cognitive functions, and which could thereby serve as proxies for the higher-level functional importance.

3.2.1 Natural Kinds

Another way of framing the purpose of this project is in terms of *natural kinds*. The overarching question then is to what extent this structural description of the graphs in terms of simplices really corresponds to any useful functional grouping. In philosophy, the notion of *natural kinds* is used to describe groupings of entities which reflect actual structures in the natural world and not just things humans group together based on their actions and motivations [6]. In other words, a grouping must be mind-independent in order to qualify as a natural kind. As an example, chemical elements are normally seen as natural kinds, whereas tables are not. In the case of simplices, one could say that the structural connectivity makes them a natural kind. However, it is not really the structure itself we are interested in, but rather the function it facilitates. The assumption is that the structures imposes certain constrains on the information-flow between the neurons in the simplices, such that it also makes sense to treat them as distinct groups based on their functional role. Whether this is the case is far from certain.

Firstly, the directed simplices are defined without regards to the nature of the neurons (whether they are excitatory or inhibitory), nor the synaptic weights in the structures. How these properties are distributed in the simplex will naturally have some impact on the functional output of the simplex, so it is far from obvious that directed simplices of the same dimension, but with very different distributions of synaptic weights and neuron types, can meaningfully be grouped together in terms of their functional potential.

Secondly, one must not forget that these simplices do not exist in isolation. Whereas it seems reasonable to treat semi-isolated clusters in networks as functional units, since they only have a few incoming and a few outgoing nodes, and therefore share much of the same information, this is not the case when it comes to simplices. The neurons which form part

of the simplices are also highly connected with neurons from outside the simplex, and there will be information flowing both in and out of the simplex in far more places than just at the source and the sink. Further, neurons have no way of distinguishing between activity coming from different sources, which further speaks against treating the simplices as clearly delineated functional units.

An underlying assumption behind the focus on simplicial structures is that despite these sources of variance, the simplices nevertheless form meaningful functional groups according to some metric. In this project, different types of groupings were tested relative to the different metrics outlined below. The most obvious grouping is that of simplex size, where the expectation seems to be that higher-order simplices are somehow more powerful or significant in performing some function in the brain. Another underlying assumption is that the complete simplices somehow represent a distinguished functional group amongst all the other (arbitrary) structures one could have chosen to focus on. In this case, simplices with missing edges lose their status as simplices, whereas additional edges, either to other neurons outside of the simplex or reciprocal connections within the simplex, are allowed. It is far from trivial why this should be the set of necessary and sufficient conditions for something to be a directed simplex. To justify having the completeness of the simplex as a necessary requirement when treating the simplicial structures as functional groups, one would need the difference between the complete and incomplete simplicial structures to be functionally observable. On the other hand, the definition of a directed simplex allows for additional connections in the graph. Thus, for the argument to hold, one must be able to show that adding additional connections has a lesser overall effect on the function than removing connections, and preferably that the complete simplex is somehow in a special functional group, worthy of being singled out. The validity of these groupings were tested using the appropriate analyses of variance, as presented in section 2.6.4.

3.2.2 Quantifying Significance - In Search of a Metric

A difficulty was thus to find a meaningful metric which could be used to quantify the significance of these structures. For example, one could consider the effect the simplicial structures have on the general information-flow in the network, or compare the information input in the source-neuron with the output of the sink-neuron. In the case of neurons firing, we know that the information is encoded in the spike trains, not in the individual action potentials. However, talk of information only makes sense once you have an encoding and a decoding protocol, and since different neurons play different roles in the network, no universal decoding scheme can be assumed, so the same patterns of spiking might mean completely different things coming from different neurons. To make matters even worse, the activity in the simulated networks studied here remains uninterpreted. The input cannot be related to any specific sensory stimuli, nor the output to any function. We simply have noise in and noise out, making it very difficult to use this dataset to assign function to structures, both at the local and the global scale.

For these reasons, coming up with any one metric which could be used to quantify the functional significance of higher-order simplices was simply impossible, so a small subset of metrics were chosen. These include the average treatment effect, transfer entropy and autocorrelation, details of which will be given in section 4. The first two metrics mostly focus on

the causal relationship between the source and the sink neuron, whereas the autocorrelation mainly focuses on the activity patterns in the sink neuron as function of individual spikes in the source neuron. Importantly, these are only a few of a near infinite number of possible functional roles these simplicial structures might have, so a negative result on these metrics is by no means sufficient to disprove the functional role of the simplicial structures in general. In addition, an alternative approach was also tested. Without making any assumptions about the actual function, it was tested whether it was possible to train a machine learning model to learn the mapping from the activity in the networks to their simplicial structures.

4 Analysis of Function

4.1 Methodological Approach

As discussed above, the general challenge when testing for the functional role of directed simplices is that the question itself is somewhat ill defined. The underlying assumption in much of the literature seems to be that it makes *intuitive* sense that such a neatly defined structure should come with a corresponding functional role, but what this role is remains unspecified. In general, when testing for function in a dataset, one must specify the hypothesis beforehand, and preferably have a new dataset for each variable under investigation. This is to avoid so-called p-hacking, which is when one tests numerous variables using the same dataset in the hope that at least *something* will turn up as statistically significant, which, as a result of the sheer numbers and the probabilistic nature of p-values, it eventually will. In other words, by performing numerous different tests on the same dataset, we increase the chances of type I errors. This is because, when using a significance level of $\alpha = 0.05$, we expect to incorrectly reject the null hypothesis in 5 % of the cases. Thus, the chance of doing this for some of the observations increases when different tests are performed on the same dataset. Thus, to somewhat mitigate the problem of incorrectly interpreting random variation as statistically significant, it was decided that if anything of substance turned out to be a statistically significant functional trait, a new dataset would have to be generated to double check that this was not simply due to random fluctuations, and hence to validate the conclusions.

In the following sections, three different metrics were used to test for function. These were average treatment effect, transfer entropy and autocorrelation. With some exceptions, the general approach was to focus on the relationship between activity in the source and sink neurons, and investigate the dependency of this on the size and completeness of the simplices. Analysis of variance was then used to test whether the grouping used, be it size or completeness, was statistically meaningful and informative given the functional trait in question. The underlying premise for even considering directed simplices an interesting structural group is that they also form an interesting functional group, which is what the analyses of variance were included to test.

4.2 Average Treatment Effect

One potentially important functional trait is the causal relationship between the source and the sink neurons in simplicial structures. In [47], they only investigated the *correlations* between the spike trains of different neurons. This is an interesting measure if what one is looking for is *cofiring* or *coactivity*. However, because of its symmetry, correlation is a poor measure of causal effect, even if one includes a time-shift. Instead, the *average treatment effect*, a metric commonly used to quantify the causal effect of treatments in randomised controlled trials in medical research, was adopted. This method requires randomised interventions, but is appropriate here since the lack of confounding variables ensures that the activity in the source neurons can be treated as randomised interventions without further ado. Here, only the causal effects of the source neurons on the sink neurons were considered.

The ATE is a measure of the average difference in outcome between subjects given treat-

ment and those not given treatment. It assumes a counterfactual understanding of causality, where causation is taken as the difference in effect when one intervention is used compared with what *would have happened* had another intervention been chosen. This is done by making the variable conditioned on an intervention, so the ATE is defined using the *do-operator* introduced in section 2.3. Given outcome Y and treatment T, where T and Y are binary values $\in \{0, 1\}$, the average treatment effect τ of T on Y is given by

$$\tau = \mathbb{E}[Y|do(T = 1)] - \mathbb{E}[Y|do(T = 0)].$$

In our case, Y represents whether the sink neuron spikes or not and T represents the source neuron's spiking or not. To study the time dependency of the causal effect of T on Y, the ATE was calculated for different relative time-shifts. Thus, the activity of the source neuron at time t was compared with the activity of the sink neuron at time $t+i$. The ATE with time-shift i of network n is given by:

$$\tau_{i,n} = \mathbb{E}[Y_{t+i,n}|do(T_{t,n} = 1)] - \mathbb{E}[Y_{t+i,n}|do(T_{t,n} = 0)]$$

Since the activity is binary, the expectation values are equal to the conditional probabilities. Consequently, we have that

$$\begin{aligned}\tau_{i,n} =& 1 \times P(Y_{t+i,n} = 1|do(T_{i,n} = 1)) + 0 \times P(Y_{t+i,n} = 0|do(T_{i,n} = 1)) \\ & - (1 \times P(Y_{t+i,n} = 1|do(T_{i,n} = 0)) + 0 \times P(Y_{t+i,n} = 0|do(T_{i,n} = 0))) \\ =& P(Y_{t+i,n} = 1|do(T_{i,n} = 1)) - P(Y_{t+i} = 1|do(T_{i,n} = 0))\end{aligned}$$

For simplicity, $do(T_t = 1)$ is denoted as T and $do(T_t = 0)$ as $\neg T$. From the definition of conditional probabilities, we have that

$$P(Y|T) = \frac{P(Y \cap T)}{P(T)}$$

and

$$P(Y|\neg T) = \frac{P(Y \cap \neg T)}{P(\neg T)}$$

Substituting in

$$P(Y) = P(Y \cap T) + P(Y \cap \neg T),$$

we have that the absolute ATE is

$$\begin{aligned}\tau_{abs} &= P(Y|T) - P(Y|\neg T) \\ &= \frac{P(Y \cap T)}{P(T)} - \frac{P(Y) - P(Y|T)}{1 - P(T)}\end{aligned} \tag{4}$$

This gives the *absolute* ATE. This is equivalent to the *effective connectivity* between the source and sink neurons in the network, which can also be defined for different time-shifts between them. Perhaps more informative is the *relative* ATE, which tells us the relative

change in activity of the sink neuron when the source neuron fires. This is given by the absolute ATE divided by the probability of the sink neuron's firing given that the source neurons did *not* fire.

$$\tau_{rel} = \frac{\tau_{abs}}{P(Y|\neg T)} \tag{5}$$

These values describe individual networks. However, the aim was to find functional properties which simplicial structures of the same dimensionality have in common. Thus, the average treatment effect across networks of the same size was calculated, such that the relative ATE for time-shift i of the N networks containing d-dimensional simplices in the dataset was given by

$$\text{ATE}_{rel,i,d} = \frac{1}{N}\sum_{n=1}^{N} \tau_{rel,i,n},$$

and similarly for the absolute ATE

$$\text{ATE}_{abs,i,d} = \frac{1}{N}\sum_{n=1}^{N} \tau_{abs,i,n}.$$

4.3 ATE in Minimal Networks

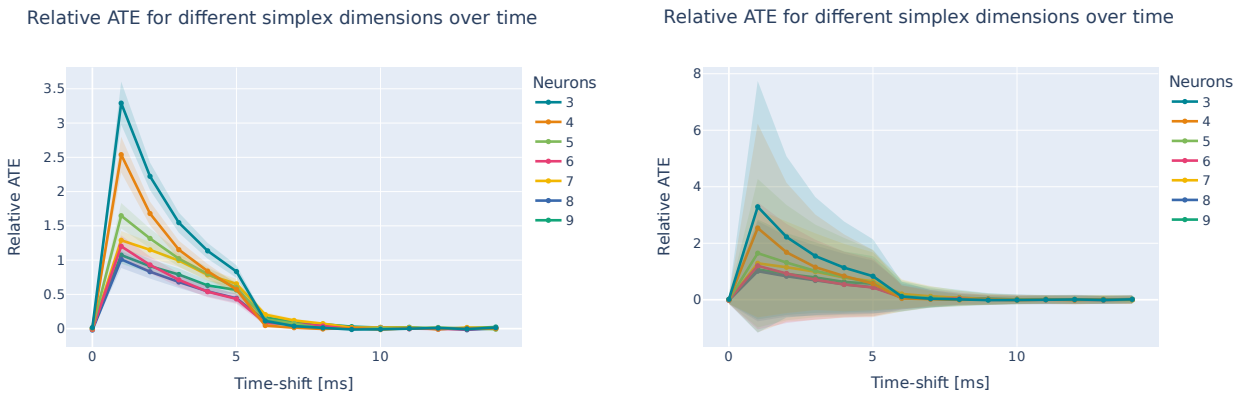
One potential functional trait associated with simplicial structures is the average treatment effect between the source and the sink neuron. This is a measure of the average change in activity in the sink neuron induced by an extra spike in the source neuron. In the idealised networks, where there are no confounding factors and the only structure is a perfect directed simplex, all the information is constrained to move from the source to the sink, so the two neurons form the two bottlenecks in the information flow. In such a case, one can treat the entire simplex as a single causal unit, such that the ATE between the source and the sink neurons also becomes a measure of the causal effect the unit as a whole exerts on the incoming information, by comparing its input with its output.

As the time-shift between the spike-trains increases, this will not just measure the direct causal effect, but also include the causal effect passing through the intermediate neurons in the simplex. It should be noted that this is a somewhat crude measure of causal effect, since, as discussed above, this only measures the increase in probability of individual spikes. In the case of neural signalling, we rarely care about such independent spikes. However, even though the actual causal effect we are interested in might be more complex, it is reasonable to assume that this would also lead to an increase in the ATE for individual spikes. Thus, the ATE should function as a reasonable proxy, given that we lack a description of the "true" function.

If the assumption is that bigger simplices are somehow more "important" than smaller simplices, and we assume that the ATE between the source and the sink neurons is a good measure of importance, then we should expect to find a higher ATE for bigger simplices. Note that the ATE is also a function of the time-shift used between the spike-trains of the two neurons. A reasonable expectation is that the causal effect of the source neuron will last

for longer for simplices of higher dimensions, since the maximal path length between them will increase for each additional neuron. This would show as a relative increase in ATE for bigger time-shifts.

4.3.1 Results



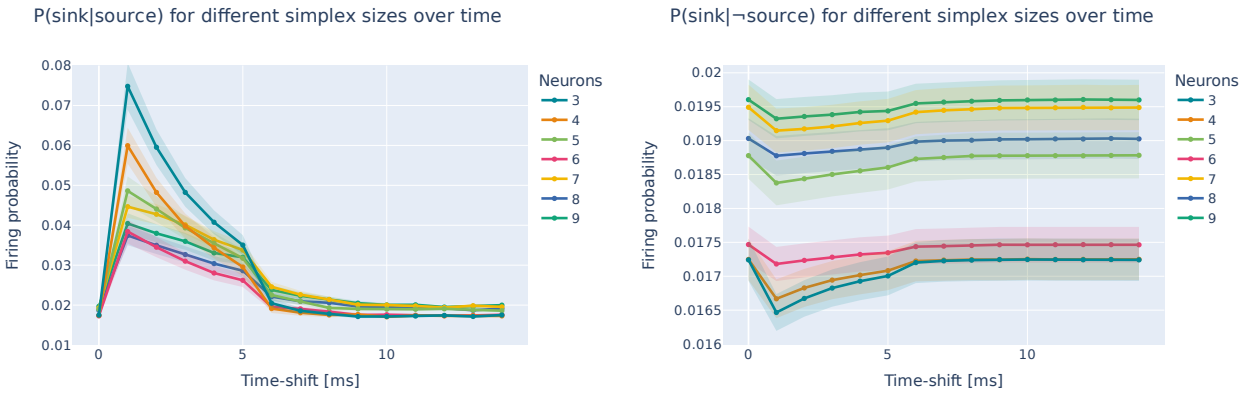
(a) Standard errors shown in shaded regions.

(b) Standard deviations shown in shaded regions.

Figure 9: The average relative ATE between source and sink neurons as function of time-shift between the time-series, shown for simplices of different sizes. Standard errors and standard deviations are shown in the shaded regions of subfigures (a) and (b) respectively.

First, the relative ATE was studied for simplices containing between 3 and 9 neurons, as shown in fig. 9, with the standard errors included in subfig. (a) and the standard deviations in subfig. (b). Here, the relative ATE is plotted as function of time-shift between the spike-trains of the source and sink neurons. There are two things worth noting here. Firstly, the relative ATE is strongest for the first time-step across all simplex sizes, suggesting that the main causal effect is transmitted through the direct connection from the source neuron to the sink neuron, regardless of simplex size. As time passes, the relative ATE decreases, until after about 6 ms, where there is hardly any detectable causal influence for any of the simplex sizes. Secondly, the relative ATE is significantly stronger for the smaller simplices, even for greater time-shifts. This is contrary to the expectation that the additional causal paths in the bigger simplices from the source to the sink would make the effect last for longer. This is not what is observed, however, and we conducted some further investigations to try to get a better idea of why this might be.

From eq. (5) and eq. (4), one can see that the relative ATE must come from a combination of the probabilities $P(sink|source)$ and $P(sink|\neg source)$, so these were investigated further. Again, this was done for different relative time-shifts between the spike-trains of the two neurons. One potential reason for the reduced relative ATE for simplices of higher dimensions is that the sink neurons in the larger simplices show less variation in their spike rate over time. This can be seen in fig. 10b, where the probabilities of the sink firing given that the source did *not* fire is plotted as function of time-shift for simplices of different sizes. The likelihood that a sink neuron fires without the source neuron's firing is higher for simplices of higher dimensions, giving a lower both absolute and relative ATE. Further, the average



(a) $P(sink|source)$ (b) $P(sink|\neg source)$

Figure 10: The average probabilities of the sink firing given that the source neuron in the simplex did (a) or did not (b) fire as function of the time-shift between the two time-series. Shown for simplices of different sizes, with the standard errors of the means shown in shaded regions. The errors might make it difficult to distinguish the colours in (b), so in order of high to low probabilities, the simplex sizes are as follows: 9, 7, 8, 5, 6, 4, 3.

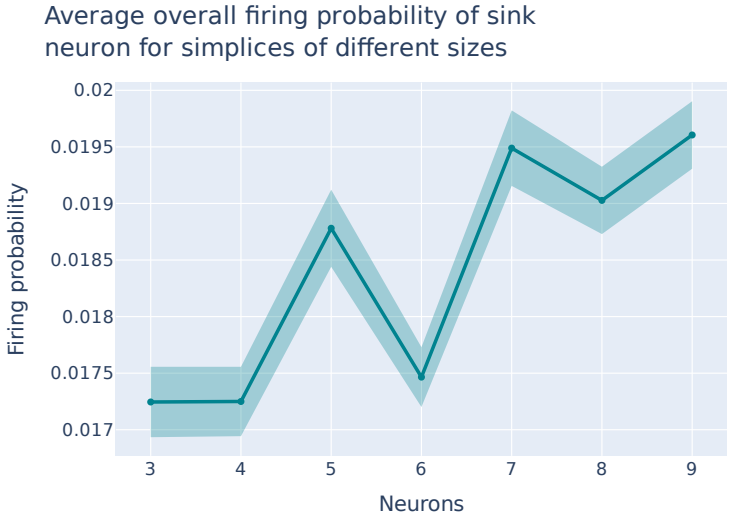
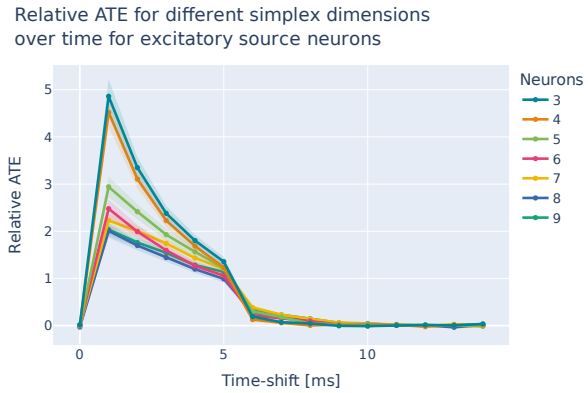
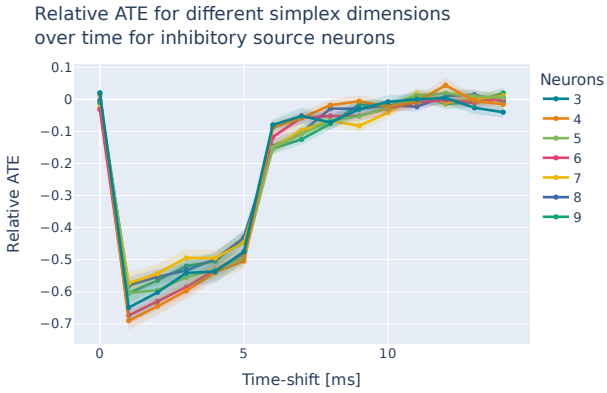


Figure 11: Average probability that the sink neuron fires at any given time step as function of simplex size. Standard errors shown in shaded region.

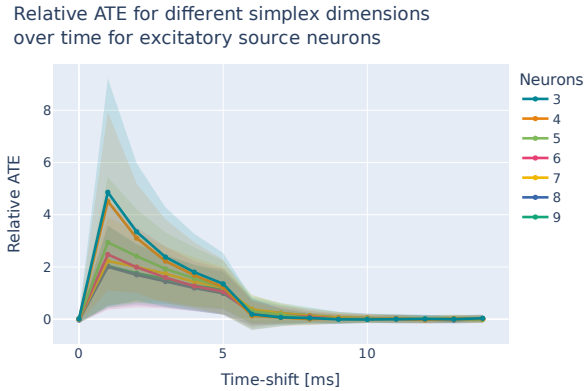
firing rate for the sink neurons also tend to increase with the simplex dimension, as shown in fig. 11, where the average overall firing probability is plotted as function of simplex size. This is likely a result of the higher number of incoming connections to these neurons compared to those in the smaller simplices, which introduces more complexity to the dynamics of the sink neurons. Note that this does not necessarily generalise to more interconnected networks, where the sink neurons of smaller simplices might be tightly interconnected with other parts of the networks, and therefore have higher in-degrees than the ones in the idealised networks considered here. In fig. 10a, the probability of the sink neuron's firing given that the source neuron fired is plotted as function of time-shift for different simplex sizes. As can be seen,



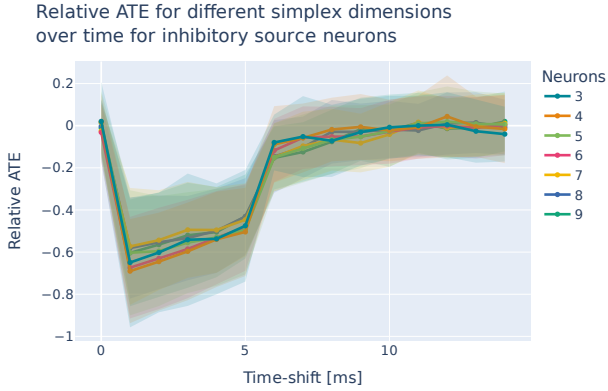
(a) Excitatory source neurons, standard errors in shaded regions.



(b) Inhibitory source neurons, standard errors in shaded regions.



(c) Excitatory source neurons, standard deviations in shaded regions.



(d) Inhibitory source neurons, standard deviations in shaded regions.

Figure 12: The average relative ATE between source and sink neurons as function of time-shift between the time-series, shown for simplices of different sizes. Standard deviations are shown as shaded areas. The dataset has been sorted based on whether the source neuron is excitatory or inhibitory, with the excitatory shown in (a)/(c) and the inhibitory shown in (b)/(d), with standard deviations included in (a)/(b), and standard errors in (c)/(d).

this follows a very similar pattern to the relative ATE in fig. 9a. The sink neurons in the smaller simplices have a significantly higher probability of firing right after the source neuron has fired. Again, this difference is somewhat surprising, since the direct connection from the source to the sink is present in all of the simplices. One potential reason for this is that as a result of the firing patterns of the higher order simplices being more spread out, the sink neurons of these simplices are more likely to be in their refractory periods when the source neuron fires next, reducing their chance of firing.

A final feature detected in these graphs is that the firing probabilities are not simply a function of the simplex size. For the simplices consisting of more than 5 neurons, both the average firing rate and the probability of firing when the source neuron did not fire is lower for the even-numbered simplices than for the odd-numbered ones. This can be seen in fig. 10 and fig. 11, where the sink neurons in the simplices with 6 and 8 neurons have

significantly lower firing probabilities than expected from the probabilities of the simplices with 5, 7 and 9 neurons (if we assume linearity). Why this is the case was not investigated further, but a potential reason is that the time-phase of the inputs somehow corresponds with the refractory period of the sink neurons.

Finally, and rather importantly, one must consider the variance in the datasets, and not just the standard errors of the means. The standard deviation is a measure of the average variance of the datapoints in the groups, so is indicative of how homogeneous the group is with respect to the metric in question. The standard deviations of the relative ATEs are shown in the shaded regions in fig. 9b. Looking at this, it is clear that the variance in the relative ATE of simplices of the same dimension is very high. Indeed, it looks as if the average values of all the sizes considered are within the standard deviations of all the other sizes. This is a strong indication that we are dealing with some very heterogeneous groups of entities, and one might begin to question whether grouping them together based on size with respect to the relative ATE is even meaningful. It is not difficult to think of reasons for the high standard deviation, for although the simplices of a given dimension all share the same structure, the distribution of synaptic weights and inhibitory and excitatory neurons can vary widely. This will certainly have a strong effect on the firing patterns of the different neurons.

One obvious factor when looking at the ATE between the source and the sink is of course whether the source neuron is excitatory or inhibitory. We therefore tested the effect of also grouping the networks together based on the type of the source neuron. The resulting graphs, again showing the relative ATE for simplices of different sizes, but this time with the networks with excitatory and inhibitory source neurons in separate groups, are shown in fig. 12. When this feature is corrected for, the standard deviations in both of the datasets become smaller, showing that this was indeed a contributing factor to the variance in the previous analysis. However, the variance is, not surprisingly, still high, since there will still be great variety in both the synaptic strengths and the distribution in type of the other neurons.

Finally, something which might be worth noticing is that the positive effect of the excitatory source neurons is far stronger than the negative effect of the inhibitory source neurons. The absolute values of the ATE of the excitatory source neurons is about 10 times higher than the absolute values of the ATE of the inhibitory source neurons. There is also a much clearer dependency on simplex size for the ATE of the excitatory source neurons than for the inhibitory ones, which are more clustered together. This suggests that the size of the simplex only makes a difference to the ATE between source and sink when the source neuron is excitatory.

Analysis of Variance Given the high in-group variance, a pressing question in this case is whether categorising the different graphs according to simplex size is even informative. To test this, the intention was to use the one-way ANOVA, so we first tested whether the observations within each group were normally distributed and whether their variance was equal.

The resulting analysis showed that the distributions were generally normally distributed for time-shifts above 10 ms, but were in general *not* normally distributed for time-shifts less

than that. This held across all simplex sizes. Unfortunately, the variances were not homogeneous for time-shifts below 5 ms. This makes employing either variation of the ANOVA test considered unreliable. We therefore tested whether grouping the simplices according to whether the source neuron was inhibitory or excitatory might help. The distributions were still not normal for lower time-shifts, nor were the variances equal. Unfortunately, none of the tests of variance were thus be applicable, as they would not give reliable results if used.

Thus, the weaker Mann-Whitney U test was used with pairwise comparisons of all the different simplex sizes for time-shifts up to 10 ms. The null hypothesis in each of these cases was that the underlying distributions behind the two datasets were identical. The null hypotheses were rejected for p-values below 0.05, meaning that the difference between the two groups was statistically significant, and thus that their difference in simplex size corresponded with a difference in relative ATE.

The results corresponded well with what can be seen from visual inspection of fig. 9 and fig. 12. When the type of source neuron was not controlled for, only the smallest simplices of 3 neurons were significantly different from most of the other simplex sizes for time-shifts up to 5 ms. The 3-simplices (4 neurons) were also significantly different from most of the other sizes for a time-shift of 1 ms. When separating the networks with excitatory and inhibitory source neurons, the significance levels of the excitatory group were higher than for the inhibitory group. They also corresponded well with the patterns observed without the pre-grouping, but with even higher significance levels. For the graphs with excitatory source neurons, both the 2- and 3-simplices were significantly different from all the other sizes for time-shifts up to 4 ms. For the inhibitory source neurons, there was more variation across the time steps, where the 3- and 5-simplices stood out for time-shifts of 2 and 3 ms and the 2- and 3- simplices for a time-shift of 6 ms. There were few general trends in the significance of the groupings for the inhibitory source neurons.

4.3.2 Discussion

There are two main finding from these results. Firstly, the relative ATE between the source and the sink neurons is significantly higher in the smaller simplices for time-shifts up to about 6 ms, after which it is roughly identical for all simplex sizes. Secondly, the high variance in the results speaks against treating directed simplices of the same dimensions as forming functional groups, and not just structural groups.

Regarding the stronger ATE between source and sink neurons in smaller simplicies, this was contrary to the expectation that the ATE would at least be higher for bigger simplices for greater time-shifts. The ATE at larger time-shifts can either be the causal effect passing through the mediator-neurons in the simplices, or from the fact that the effect of even a direct connection works over 10 time-steps with decreasing effect (see eq. (3)). The lasting effect for the smaller simplices must indeed come from this coupling window, since there is no other way for the effect of the source on the sink in a 2-simplex (3 neurons) to last for more than 2 ms. The fact that the ATE in the smaller simplices is so much higher than for the bigger ones suggests that most of the detectable ATE comes from these direct connections, rather than from the effect passing through mediator-neurons in the bigger simplices. In this context, this is a negative result regarding the importance of higher-order simplices, though with significant limitations. One must remember that these results are

based on very idealised graphs, where the sink neurons in the smaller graphs receive far less input than those in the bigger simplices. This is not necessarily the case in more complex networks, so the difference detected could also partly be the result of a simple difference in in-degree. Yet, one would expect the two values to be correlated also in more integrated networks.

Perhaps more damning for the focus on simplices is the significant levels of variance within the groups. Although the mean values between the groups are significantly different, the within-group variance makes it difficult to justify using these groupings as strongly informative of ATE. When we search for functional groupings, we usually look for the ones that give a combination of differentiable means and low within-group variance. Assuming the means are different, high in-group variance could be an indication that the grouping we are currently working with partly corresponds with a more "natural" grouping, which is the real cause of the difference in function detected. Further, as mentioned, it is far from difficult to come up with reasons for the high in-group variance. For example, these results clearly show that whether the source neuron is inhibitory or excitatory has a strong effect on the mutual dynamics of the source and the sink neurons in the simplices, and it is far from obvious why grouping them together simply based on their simplicial structure is functionally meaningful. Further, this is only one of the differences which could be controlled for. It is likely that other factors, such as the type of some of the other neurons and the edge weights, give rise to different dynamics. This is reflected in the great variance in the relative ATE, and gives reason to be sceptical of the meaningfulness of treating all these very heterogeneous structures (i.e. simplices of the same dimension) as part of a single functional group. This point is further reinforced by the finding that there is hardly any difference between the simplices of different dimensions for networks with inhibitory source neurons, so with respect to relative ATE between source and sink neurons, distinguishing based on simplex size only seems meaningful in the cases where the source neurons are excitatory.

4.4 ATE in Imperfect Simplices

Table 4: The number of neurons and edges in complete simplices of different dimensions, as well as the number of edges to be added or removed.

Neurons	Edges	10 %	15 %
4	6	1	1
5	10	1	2
6	15	2	2
7	21	2	3
8	28	3	4
9	36	4	5
10	45	5	7

Another approach in determining the functional role of simplices is to compare the functionality of complete simplices with slightly imperfect simplices to see whether the perfect simplices somehow stand out. This would be an argument in favour of treating them as a unique functional group when compared with similar structures with a few edges added or removed. This is important because of the very precise way the simplices are defined, where additional edges are permitted, whereas the simplex loses its status as a simplex if even a single edge is removed. Assuming that this is a meaningful set of necessary and sufficient conditions for something to be a simplex, this distinction should be reflected in the functionality of these different structures.

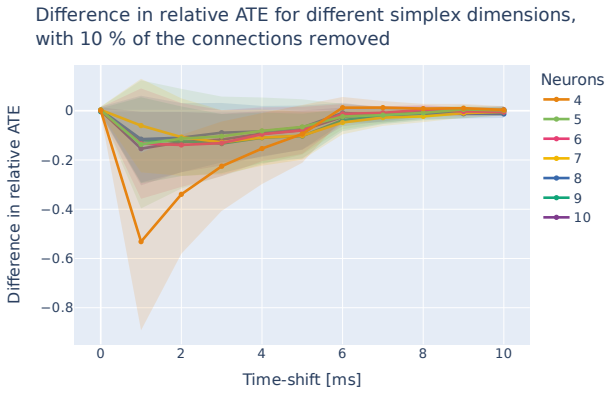
The aim here was thus to test whether there was an observable difference in relative ATE between these structures. Directed simplices of different sizes were again used, and the results in the complete simplices were compared with those in the simplices where some edges had been added or removed. A similar analysis as was used for the idealised simplices was thus performed on simplices where a random selection of 10 % and 15 % of the edges in the simplices had been added or removed (rounding off to the nearest integer). The number of edges added and removed for each simplex size is shown in table 4. It should be noted that adding extra edges to an already complete simplex when this constitutes the entire network means adding reciprocal connections between the neurons. Thus, the causal flow in the simplex is no longer uni-directional, as it will now include causal loops.

4.4.1 Results

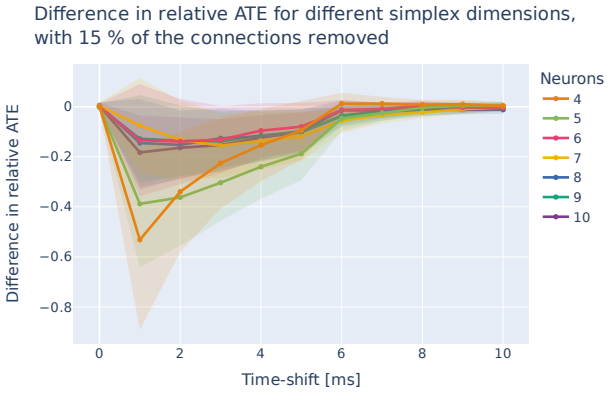
To study the effect of changing the number of edges in the simplices, the difference in relative ATE between the complete simplex and the adjusted ones were considered. These were calculated by taking the average relative ATE of the adjusted simplices minus the average relative ATE of the complete simplices, such that the effect of the change is what is plotted. The effects of removing or adding 10 or 15 % of the edges are shown in fig. 13.

As expected, the relative ATE decreases when connections are removed, and slightly more so when 15 % are removed than when 10 % are removed. The effect is most notable up to a time-shift of 6 ms, across all sizes. This corresponds with the duration of the relative ATE found in the perfect simplices, after which it is close to 0, so it is perfectly reasonable that removing edges has little effect after this. Interestingly, when 10 % are removed, the effect on the smallest simplex (4 neurons) is disproportionately higher compared to all the others. In fact, the effects on the other networks are very similar. The same pattern is found when looking at the 15 % reduction in edges, though in this case the 4-simplex (5 neurons) is also more strongly affected. This suggests that for each percentage-wise reduction, above a certain threshold in simplex-size, the effect of removing the edges is roughly the same for all simplex sizes.

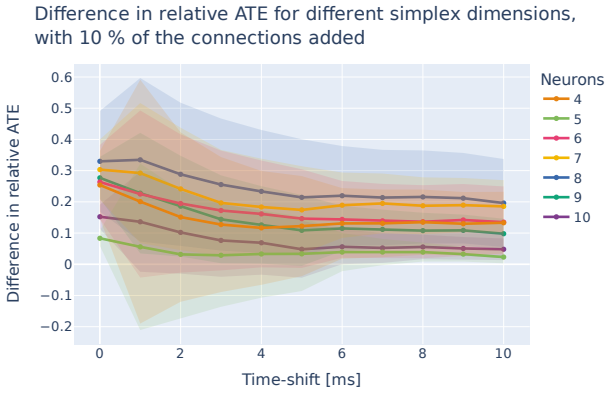
Importantly, however, the effect of *adding* edges leads to a corresponding increase in the ATE. With the exception of the smallest simplices (3 and 4 neurons), for smaller time-shifts, the absolute value of the difference in ATE from adding connections is actually greater than the effect of removing edges. The main difference is that the shapes of the curves are more similar across all simplex sizes, without any disproportionate effect for the smallest simplices. Further, the effect of adding edges is visible for longer time-shifts as well, probably because this has the effect of introducing causal loops in the networks.



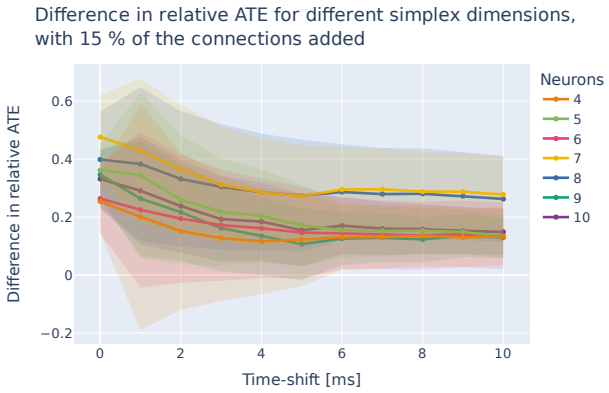
(a) 10 % decrease in edges.



(b) 15 % decrease in edges.

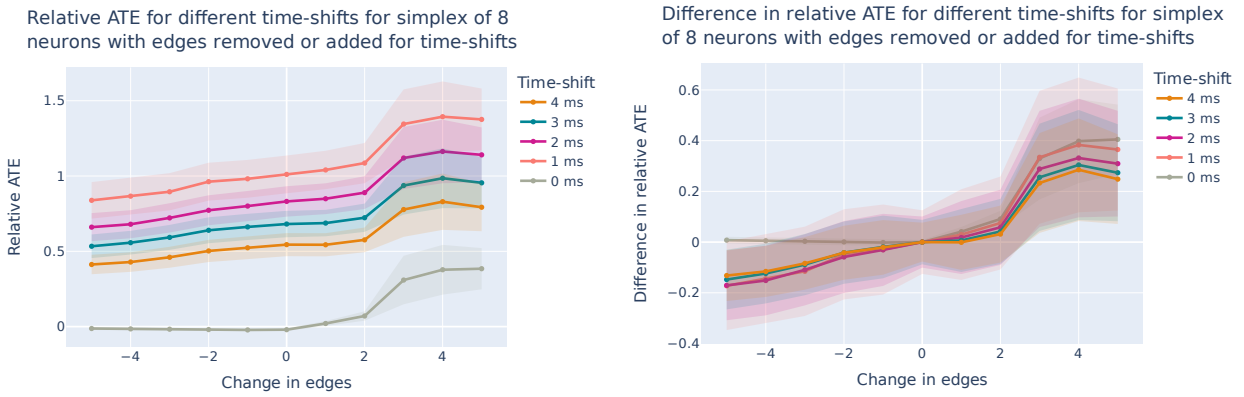


(c) 10 % increase in edges.



(d) 15 % increase in edges.

Figure 13: The difference in average relative ATE between the source and sink neurons for simplices where 10 (a)/(c) and 15 (b)/(d) % of the edges has been removed (a)/(b) or added (c)/(d) and the corresponding complete simplices. Plotted as function of the time-shift between the two time series for simplices of different sizes, with standard errors included in the shaded regions.



(a) Relative ATE, standard error in shaded regions.

(b) Difference in relative ATE, standard error in shaded regions

Figure 14: Average relative ATE between source and sink neurons in 7-simplices as function of the number of edges added or removed. (a) simply shows the relative ATE for each time-shift, whereas (b) shows the difference between the changed graphs and the complete simplex. Hence, all the lines meet in origo. Standard errors are shown in the shaded regions.

In one sense, it is a positive result that both adding and removing edges has an effect on the ATE in opposite directions. However, this equally well shows that the ATE as a function of the number of connections in the network lies on a continuum, which somewhat goes against the categorical way in which they are defined, where the complete simplex without any additional connections is treated as the special case.

To more rigorously test whether the complete simplex stands out in any way, or whether the relative ATE is fairly linear with respect to the number of edges in the graph, the effect of adding and removing edges incrementally in a 7-simplex was tested. The change in the number of edges was on the interval $[-5, 5]$, and the difference in the relative ATE between the source and the sink for time-shifts up to 4 ms was plotted as function of the changing number of edges, show in fig. 14. From these results, there is nothing which stands out with the complete simplex without additional reciprocal edges. Rather, the relationship between the relative ATE and the change in edges is fairly linear for all time-shifts, up to the point where 3 edges are added. At the addition of 3 edges, there is a sharp discontinuity, where the relative ATE makes a sudden jump for all time-shifts. Interestingly, at this point even the ATE without a time-shift, which has been 0 thus far, also becomes positive. This is a strong indication that the sudden increase is a result of the reciprocal connections creating causal loops in some of the networks, which could also be the reason for the sudden increase in the standard error at this point. Perhaps at a certain number of additional edges, the likelihood of causal loops appearing reaches some threshold, creating the sudden jump in the average difference in the relative ATE. How many edges are needed to reach this threshold is likely dependent on the size of the simplex, but further investigations of this is beyond the scope of this project.

Analysis of Variance The grouping in question here is no longer the size of the simplices, but rather their completeness. More specifically, we wish to test whether the structural defi-

nition, where simplices with additional edges are grouped together with the perfect simplices whereas simplices with missing edges are not, also makes sense with respect to their assumed functionality. Since the distributions were neither normally distributed nor showed equality of variance, we used the Mann-Whitney U test to check whether there were statistically significant differences between the pairwise distributions of the three groups in question, namely the perfect simplices, the simplices with 10 and 15 % of the edges removed and the simplicies with 10 and 15 % of the edges added. This was tested separately for simplices of different sizes. A result in favour of the current definition of simplices would be that the simplices with edges removed are statistically different from both the perfect simplices and those with edges added, whereas the perfect simplices and those with edges added are *not* significantly different.

Somewhat surprisingly, there were no statistically significant differences detected between the complete simplices and those with edges removed for any of the time-shifts nor for any of the simplex sizes. This is likely a results of the high variance in the datasets, as seen from the large standard deviations. The only statistically significant differences between the distributions were found at a time-shift of 0 ms, where the datasets with edges added were significantly different from both the complete simplices and those with edges removed. This held across all simplex sizes, and is likely the result of the causal loops introduced from the additional connections.

Next, the 7-simplices with iterative changes in edges were considered. Here, all the networks with different number of edges (from 1 to 5) removed were grouped together, and so were those with edges added. These two groups were then compared with each other, as well as with the perfect 7-simplices. In this case, a statistically significant difference between the group where edges were added and the group where edges had been removed was found for time-shifts of 1 and 2 ms as well, but there was no significant difference between either of these groups and the perfect simplices.

4.4.2 Discussion

The above results clearly suggest that the number of edges between the source and the sink neuron has some effect on the relative ATE between the source and sink neurons for simplices of different sizes, although the within-group variance is too big for this difference to be statistically significant. Above a certain simplex size, the effect of removing the same percentage of connections is roughly the same for simplices of all sizes. Similarly, adding a similar percentage of edges to the simplices, giving reciprocal connections, on average increases the relative ATE. It should be noted that the type of the neurons from which the outgoing edges were either added or removed were not taken into account here. In reality, one would of course expect that whether these neurons are excitatory or inhibitory will be significant, and grouping these cases together is quite plausibly a major cause of the high variance in these datasets. Yet again, this is not something which is taken into consideration in the definition of the directed simplices, although one can easily argue that it should be. The results regarding the effect of incrementally adding or removing edges certainly speaks against giving the complete simplex any special functional status, at least with regards to the relative ATE between the source and the sink.

Finally, the fact that the effect of removing edges is lower in the higher order simplices

could be an indication that these structures are more resilient to changes in the networks, which is generally seen as a favourable trait. However, the same argument can be used in favour of granting simplices with additional reciprocal connections the same special status, since these are seemingly even more resilient to changes.

In sum, these results are insufficient to justify the strict structural definition of the directed simplices if this is thought to correspond to any function which might be measured using the relative ATE between the source and the sink neurons in simplices. The high within-group variance in all of these datasets shows that the defined groups are far from homogeneous with respect to their ATE, and there is so much overlap between the different datasets that the groupings are not even statistically significant. Thus, if the function of interest is the relative ATE, this provides a strong argument against granting complete directed simplices any special status, as there are many other similar structures which seem to serve a similar function, although these do not meet the structural criteria. However, the relative ATE between the source and the sink is only one metric out of many, and since it is not clear what function we are expecting to be associated with these simplicial structures, it is far from clear whether this is even a meaningful metric for comparing and evaluating their status as a functional group.

4.5 Transfer Entropy

Another metric which was introduced by Schreiber [48] to measure the amount of information transferred from one variable to another is that of *transfer entropy*. Under certain interpretations, this can then also be considered as a measure of the causal effect of the first variable on the second. Contrary to the ATE, this does not assume interventional data, so can also be used when only observational data is available. It is different from other metrics used to quantify causality, such as the Granger causality, in that it is also able to pick up on non-linear causal relationships between variables.

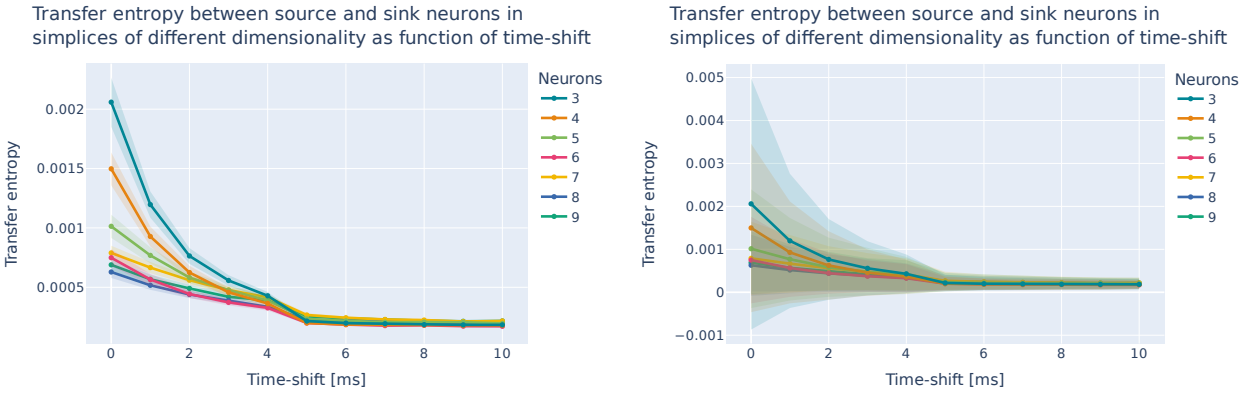
We have two variables, X and Y, where the state of X/Y at time i is denoted x_i/y_i. In our case, the variable X is the source neuron and the variable Y the sink neuron. We wish to study the causal effect of variable X on variable Y. Further, we define $\square = \{W_1, \ldots, W_l\}$ to be the background variables conditioned on. Note that no background conditions were included in our analyses. The notion of *k-histories* is also needed here. This is simply subsections of the previous history of the state of the variable, such that $y_i^{(k)} = \{y_{i-k+1}, y_{i-k+2}, \ldots, y_i\}$. Ideally, one has $k \to \infty$, but due to computational limitations, $k = 20$ was used in all analyses of the transfer entropy. The time-local transfer entropy is given by

$$t_{X\to Y,\square,i}(k) = \log_2 \frac{P\left(y_{i+1}, x_i \mid y_i^{(k)}, W_{\{1,i\}}, \ldots, W_{\{l,i\}}\right)}{P\left(y_{i+1} \mid y_i^{(k)}, W_{\{1,i\}}, \ldots, W_{\{l,i\}}\right) P\left(x_i \mid y_i^{(k)}, W_{\{1,i\}}, \ldots, W_{\{l,i\}}\right)} \tag{6}$$

and the average over time by

$$T_{X\to Y,\square}(k) = \left\langle t_{X\to Y,\square,i}(k) \right\rangle_i .$$

In practice, this was calculated using the `transfer_entropy` function of the `PyInform` package. The crucial difference between the ATE and the transfer entropy is that the ATE



(a) Standard errors in shaded regions.

(b) Standard deviations in shaded region.

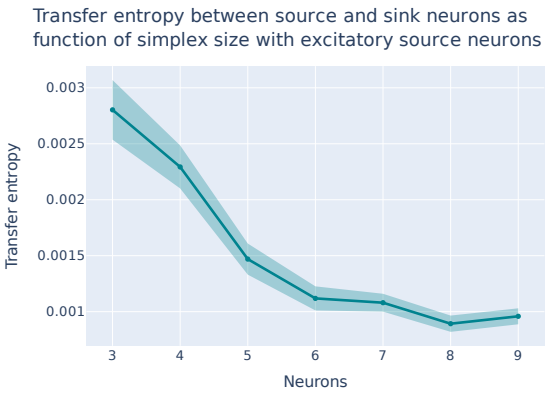
Figure 15: The average transfer entropy between the source and sink neurons in simplices of different sizes as function of the time-shift between the two time-series.

looks at events in the form of individual spikes, so only considers the likelihood that a single spike event will occur at the sink neuron a certain number of time-steps after a single spike event occurred in the source neuron. The transfer entropy, on the other hand, looks at the *k-series* of events in the sink neuron. Given the previous activity in the sink neuron over k time-steps, it quantifies the reduction in our uncertainty about its future activity from our knowing that a single spike in the source neuron has occurred. This nuance is valuable since repeating patterns are what we are truly interested in when studying neural activity, not individual spikes in isolation. Also worth noting is that whereas the ATE can be negative, the transfer entropy is always positive.

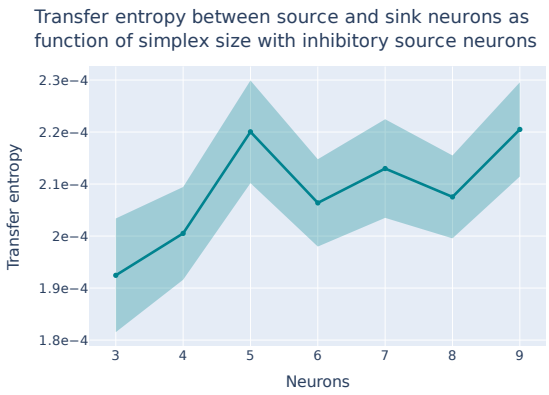
As above, this was studied as a function of of simplex size in the idealised networks and as function of the number of edges added or removed in 7-simplices. Further, since the transfer entropy can also be used in integrated networks with confounding variables, simplices within bigger networks were also investigated. Specifically, we investigated whether the sum of the outgoing and incoming transfer entropy of neurons in small-world networks was more correlated with their simplicial roles than with their in- and out-degrees.

4.5.1 Results

In fig. 15, the average transfer entropy between the source and sink neurons in simplices of different sizes is plotted as function of the time-shift between the neurons. The standard errors and standard deviations are included in the shaded regions in fig. 15a and fig. 15b respectively. The results correspond well with what was found for the relative ATE, where the transfer entropy is highest for the smallest simplices and at the first time-steps. It is worth remarking that whereas the relative ATE was 0 for time-shifts of 0 ms, the transfer entropy has its maximum at 0 ms for all simplex sizes. This is because a relative time-shift of 1 time-step is implicit in the definition, as can be seen from eq. (6), so we always consider what the current state of X tells us about the future state of Y. After about 5 ms, there is hardly any observable transfer entropy between the neurons, regardless of dimensionality of the simplices. Again, the standard deviations are remarkably high, and only the means of



(a) Excitatory source neurons



(b) Inhibitory source neurons

Figure 16: The average transfer entropy between source and sink neurons as function of simplex size for a time-shift of 0 ms. The datasets have been sorted based on whether the source neurons were excitatory or inhibitory, with the excitatory shown in (a) and the inhibitory shown in (b). Standard errors shown in shaded regions.

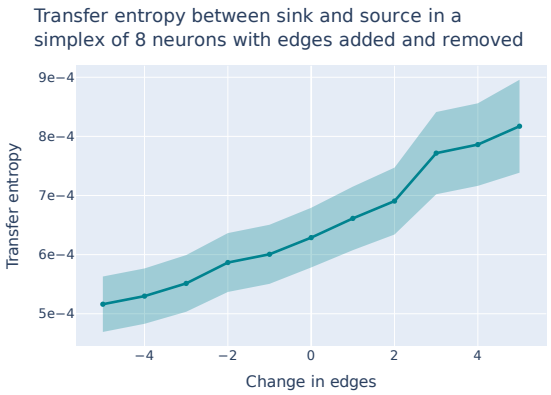
the simplices of 5 neurons or less are substantially different from the others.

As with the relative ATE, to get a better idea of the source of the variance, the type of source neurons were selected for, and the transfer entropy with a time-shift of 0 ms was plotted separately for the inhibitory and excitatory source neurons. This is shown in fig. 16. As expected, the transfer entropy is higher for the excitatory source neurons, and the same pattern of decreasing transfer entropy with increasing simplex size is found. As with the relative ATE, the absolute value of the metric for the inhibitory source neurons is about a tenth of that of the excitatory ones, showing that the effect of excitatory neurons are stronger than that of inhibitory ones. As with the relative ATE, the transfer entropy of the inhibitory source neurons do not show the same dependency on the simplex size as the excitatory ones do. Why this is so is not clear, and it should be noted that the values are extremely low, so one might wonder whether these are simply statistical fluctuations rather than actual effects, especially given the high errors.

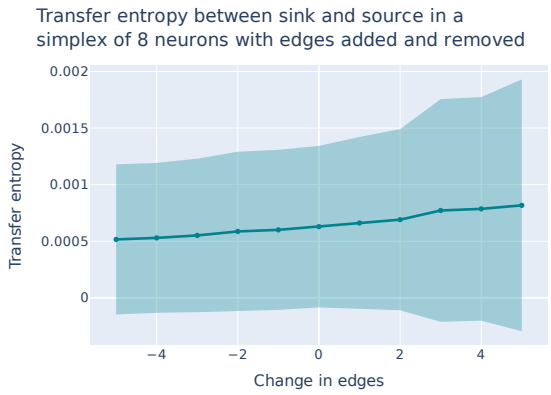
Finally, to study whether the strict structural criteria for categorising something as a simplex are justified when considering transfer entropy as the function in question, the transfer entropy as function of the number of edges added and removed in 7-simplices was plotted, as shown in fig. 17. Although the variance is high, there is a clear linear dependency between the number of edges and the transfer entropy between the source and the sink, as was found when performing the same analysis with respect to the relative ATE. Again, the standard deviations are higher for the simplices with additional edges.

Analysis of Variance Again, we tested the validity of two different types of groupings of the simplices. Firstly, the simplex size, and secondly, the number of edges added and removed in 7-simplices.

We started by testing whether the transfer entropy within each group was normally distributed for each time-shift. This assumption turned out not to hold, for neither simplex size nor for either time-shift. It was then tested whether the variance was equal between the



(a) Standard error in shaded regions.



(b) Standard deviation in shaded region.

Figure 17: The average transfer entropy for time-shift 0 ms between the source and sink neurons in a 7-simplex as a function of the number of edges added or removed.

groups for different time-shifts. For simplex sizes between 3 and 10 neurons, the variance was equal for time-shifts above 5 ms, but unequal for time-shifts smaller than that. Looking at fig. 15, it is clear that the variance is much greater for the smaller simplices for the smaller time-shifts, so this is not too surprising. Thus, the Mann-Whitney U test was again used for pairwise comparisons.

First, it was used to test whether grouping the simplices by size was informative when considering the transfer entropy between the source and the sink neurons for different time-shifts. For the early time-shifts, it was for some reason the simplices with 6 neurons which stood out most strongly from the others over time, whereas the 2-simplices were only significantly different from the others for the 0 ms time-shift. This is somewhat surprising when looking at fig. 15a, but is likely caused by the fact that the variance is much higher for the smaller simplices for the early time-shifts, as can be seen in fig. 15b. Thus, the difference in the mean is not sufficient for it to be statistically significant. When considering the networks with excitatory and inhibitory source neurons separately and for a time-shift of 0 ms, the 2- and 3-simplices were significantly different from all the other simplices (though not from each other) for the excitatory networks, whereas only the 2- and the 8-simplices were significantly different from each other in the inhibitory networks.

To test whether grouping the simplices with edges added and removed separately is meaningful, the 7-simplices with edges added and removed were considered. Again, all the networks with different number of edges removed were grouped together, and so were those with edges added, and both groups were compared with the complete simplices as well as with each other. In this case, there was a statistically significant difference between the group with edges added and the one with edges removed for all the time-shifts considered (up to 9 ms). More importantly, there was also a statistically significant difference between the complete simplices and those with edges removed for time-shifts up to 4 ms, but no corresponding difference between the group with edges added and the complete simplices. This was also tested for each incremental change separately, and the same result was found. Further, it was tested with the networks of different sizes where 10 and 15 % of the edges had been added or removed, and for most of the networks, there was a significant difference

between the networks with edges removed and both the perfect simplices and those with edges added. No significant difference between the perfect simplices and those with edges added detected was detected for any of the datasets.

Table 5: Correlation coefficients and their corresponding 95 % confidence intervals, first variable in the columns, second variable in rows, for small-world networks of 20 neurons, with time-shifts of 0, 5, and 9 ms.

0 ms					
	Incoming TE	CI95%		Outgoing TE	CI95%
In-degree	0.320833	[0.29, 0.35]	Out-degree	0.204819	[0.17, 0.23]
2-sink	0.256591	[0.23, 0.29]	2-source	0.156361	[0.13, 0.19]
3-sink	0.093902	[0.06, 0.12]	3-source	0.05115	[0.02, 0.08]
4-sink	-0.001867	[-0.03, 0.03]	4-source	-0.000091	[-0.03, 0.03]
5 ms					
	Incoming TE	CI95%		Outgoing TE	CI95%
In-degree	0.21889	[0.19, 0.25]	Out-degree	-0.012729	[-0.04, 0.02]
2-sink	0.174971	[0.14, 0.2]	2-source	-0.001758	[-0.03, 0.03]
3-sink	0.058583	[0.03, 0.09]	3-source	-0.004001	[-0.03, 0.03]
4-sink	-0.002581	[-0.03, 0.03]	4-source	0.004179	[-0.03, 0.04]
9 ms					
	Incoming TE	CI95%		Outgoing TE	CI95%
In-degree	0.224829	[0.2, 0.25]	Out-degree	-0.009327	[-0.04, 0.02]
2-sink	0.185185	[0.16, 0.21]	2-source	-0.003816	[-0.03, 0.03]
3-sink	0.074686	[0.04, 0.11]	3-source	0.000724	[-0.03, 0.03]
4-sink	0.002759	[-0.03, 0.03]	4-source	-0.000412	[-0.03, 0.03]

Small World One benefit of the transfer entropy as compared with the ATE is that it can also be used for observational data with confounding variables. From the analysis of the isolated simplices, it seems not to be the case that increasing the simplex size increases the transfer entropy between the source and the sink neurons. However, it could still be the case that the simplicial role of a neuron is connected with a more overarching "causal role" of the neuron in the network. To measure this, the transfer entropy between every pair of neurons in small-world networks was measured for different time-shifts, and the sum of the transfer entropy of a neuron to all the other neurons in the network was taken as a measure of its causal influence in the network. Similarly, the sum of the transfer entropy from all the other neurons *to* a given neuron was taken as a measure of how "influenced" the neuron is by the other neurons in the network. In other words, the outgoing transfer entropy was taken as a measure of the effect each neuron had on the dynamics of the other neurons in the network and the incoming transfer entropy as the effect the other neurons had on this neuron's dynamics.

Table 6: Correlation coefficients and their corresponding 95 % confidence intervals, first variable in the columns, second variable in rows, for small-world networks of 50 neurons, with time-shifts of 0, 5, and 9 ms.

0 ms					
	Incoming TE	CI95%		Outgoing TE	CI95%
In-degree	0.227134	[0.21, 0.25]	Out-degree	0.120497	[0.1, 0.14]
2-sink	0.146584	[0.13, 0.17]	2-source	0.070648	[0.05, 0.09]
3-sink	0.085665	[0.07, 0.11]	3-source	0.035626	[0.02, 0.06]
4-sink	0.031307	[0.01, 0.05]	4-source	0.009238	[-0.01, 0.03]
5 ms					
	Incoming TE	CI95%		Outgoing TE	CI95%
In-degree	0.16411	[0.14, 0.18]	Out-degree	-0.033858	[-0.05, -0.01]
2-sink	0.098669	[0.08, 0.12]	2-source	-0.033653	[-0.05, -0.01]
3-sink	0.054677	[0.04, 0.07]	3-source	-0.023216	[-0.04, 0.0]
4-sink	0.016058	[0.0, 0.04]	4-source	-0.009392	[-0.03, 0.01]
9 ms					
	Incoming TE	CI95%		Outgoing TE	CI95%
In-degree	0.172499	[0.15, 0.19]	Out-degree	-0.038726	[-0.06, -0.02]
2-sink	0.107778	[0.09, 0.13]	2-source	-0.040258	[-0.06, -0.02]
3-sink	0.061283	[0.04, 0.08]	3-source	-0.028408	[-0.05, -0.01]
4-sink	0.019223	[0.0, 0.04]	4-source	-0.008752	[-0.03, 0.01]

The question was then whether these properties, denoted as the outgoing and incoming transfer entropy (TE) respectively are correlated with the simplicial role of the neurons. Counts of how many 2, 3 and 4-simplices each neuron was the source and sink of was used. In addition, the in- and out-degree of each neuron was recorded. This is simply a count of the number of incoming and outgoing synapses each neuron has. The reason for including this is that one would expect the simplicial role of each neuron to be indicative of its causal role just from the fact that this is correlated with the number of incoming and outgoing synapses. Thus, if the transfer entropy was more strongly correlated with a neuron's role in bigger simplices than with its in- and out-degree, this would provide an argument in favour of treating higher-order simplices as important causal units.

Using the `pingouin` package in python, the correlation coefficients between pairs of these variables were found. The values found for networks of 20 and 50 neurons with time-shifts of 0, 5 and 9 ms are shown in table 5 and table 6 respectively. Interestingly, the correlation between both in-degrees (and sink-count) and the "incoming" transfer entropy is much stronger than the correlations between the out-degrees (and source-counts) and the "outgoing" transfer entropy. The reasons for this were not investigated further. Most important for the current investigation is the fact that the variable most strongly correlated with the transfer-entropy of a neuron is its degree, and *not* its simplicial role. This holds for both

network sizes and for all the time-shifts considered. Secondly, the correlation between the transfer entropy and the source/sink-count drops with increasing sizes of simplices. This result goes straight against the main premise of this project, which is that the number of higher dimensional simplices of which a neurons is a source and/or sink is correlated with that neuron's causal role in the network.

4.5.2 Discussion

From the investigations of the dependency of transfer entropy between source and sink neurons on the simplex size in perfect simplices, it was found that the larger the simplices, the smaller the transfer entropy, across all time-shifts. This goes against the assumption that larger simplices are somehow more causally significant than smaller ones. However, it must be noted that the variance in all the datasets was very high, especially for the smaller time-shifts, so the conclusion from the analysis of variance was that the grouping according to simplex size was in general not statistically significant for any of the pairs. Thus, it must therefore be considered a null-result with regards to the dependency of transfer entropy on simplex size.

The main result from this investigation is probably the finding that the incomplete simplices clearly stand out from both the perfect ones and the ones with added edges, whereas the perfect ones and those with added edges are not significantly different according to the Mann-Whitney U test. This was somewhat surprising, especially considering the apparent linearity of the relationship as shown in fig. 17. It thus remains unclear exactly *how* the distributions differ, and thus how the imperfect simplices are significantly different from the other two groups. However, it is certainly an indication that the completeness of the simplices has *some* significance for the causal dependencies between the source and the sink neurons in these structures. Further investigations would be needed to study the nature of this dependency more in depth.

Finally, the hypothesis that the causal role of the neurons in small-world networks is more strongly correlated with their role in higher-order simplices than with their in- and out-degree was clearly falsified. From these investigations, the main factor (amongst those here considered) in determining the causal role of a neuron in an integrated small-world network is clearly its degree, both incoming and outgoing.

In sum, these results are insufficient to justify treating transfer entropy between source and sink neurons as a function attributable to higher-order simplices.

4.6 Autocorrelation

As discussed, the information in the spike-trains of a neuron is encoded both in the firing rates and the firing intervals. When looking at the ATE and the transfer entropy, only the instantaneous firing rates of the sink neuron as function of individual spikes in the source neuron were considered. However, another way in which the simplicial structures might be significant is by facilitating certain spiking patterns. This is more in line with the standard approaches for measuring the functional role of network motifs in neuroscience. A common approach for such analyses is to look for repeating patterns in the spike trains of neurons, either individual neurons or groups thereof. As an example, Curto and Morrison showed

how different small circuit motifs could work as stable attractors of the dynamics of their constituent neurons [39]. In our case, we shall limit ourselves to studying the complexity of the activity in the sink neurons.

Repeated patterns in the time-series can be detected by measuring the autocorrelation of a time-series at different intervals. This can also be treated as a proxy for the so-called signal-to-noise ratio, which is the ratio between the desired signal and the background noise in the spike-trains of the neurons. The higher this is, the more reliable is the process by which it is generated. By comparing the autocorrelation of the activity in the sink neurons of different simplices, which is related to the signal-to-noise ratio of their output, this tells us something about the relative reliability of these structures in converting the activity in the source neuron into informative activity in the sink neuron.

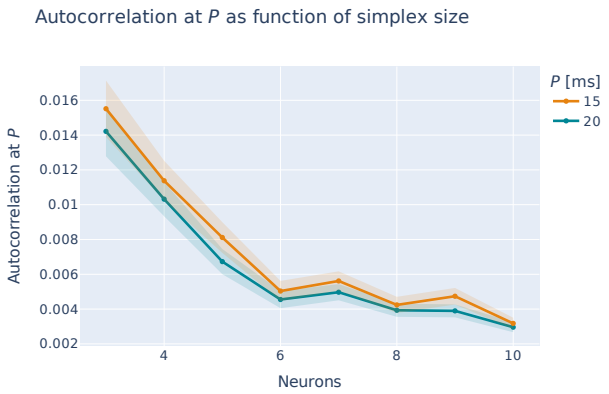
In the datasets studied thus far, the activity in the source neuron has been fairly randomly distributed. If the assumption is that each spike in the source neuron causes a certain spiking pattern in the sink neuron, we require a regular spiking pattern in the source neuron in order for this to be detectable using auto-correlation. Further, from a causal inference point of view, this gives interventional data, which provides much stronger grounding for inferring causal relations between the variables. Thus, a stimulation of the source neuron was included in the simulation of activity in the idealised simplicial networks as well as in the 7-simplices with edges added and removed. This stimulation consisted of a single spike in the source neuron with a constant period of P ms throughout the entire simulation. Note that this method does not assume that the resulting activity pattern is the same across simplices of the same dimension, only that the level of autocorrelation for the individual time series is correlated with the size or completeness of the simplices.

The autocorrelation was calculated by measuring the correlation between the initial time-series of the sink neuron with the same time-series with an added time-shift. In practice, this was done using the `acf` method from `statsmodels`, which calculated the autocorrelation for every time-shift up to a set limit, here set to 50 ms. P of 20 and 15 ms were tested. To prevent the activity in the network from exploding, the spiking threshold θ was then increased to 4.5 and 4.7 respectively.

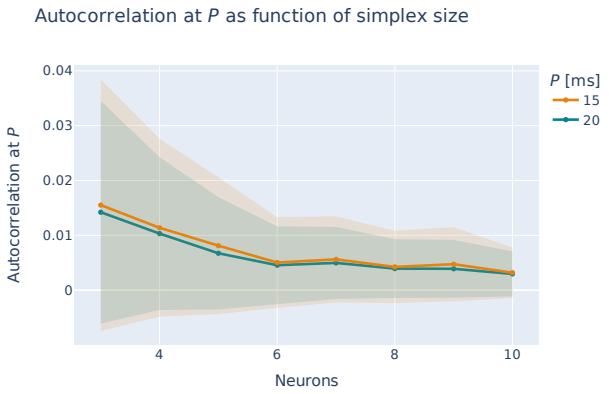
We expected to see a peak in the autocorrelation in the spike-trains of the sink neurons at P, as well as at integer multiples thereof. The main question was then whether the size and completeness of the simplex had any effect on either of two variables, the first being the strength of the autocorrelation at time-shift P, the second the distribution of autocorrelations for other time-shifts. For example, one could imagine that bigger simplices also show autocorrelation for longer time-shifts, since the influence from the source neuron is expected to last for longer. This would be indicative of longer sections of repeating patterns in the spike trains of the sink neurons, which in turn have the potential to encode more complex informational patterns.

4.6.1 Results

The average maximum autocorrelation was plotted for complete simplices of different sizes for $P = 15$ ms and $P = 20$ ms, as shown in fig. 18. As can be seen, the autocorrelation at P decreases significantly with the size of the simplices. The average autocorrelation at $P = 20$ ms for the 9-simplices is ¼ of that of the 2-simplices. However, there is also less variance

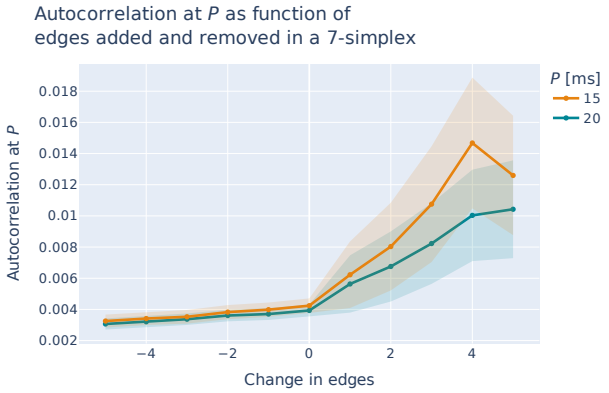


(a) Standard errors shown in shaded regions.

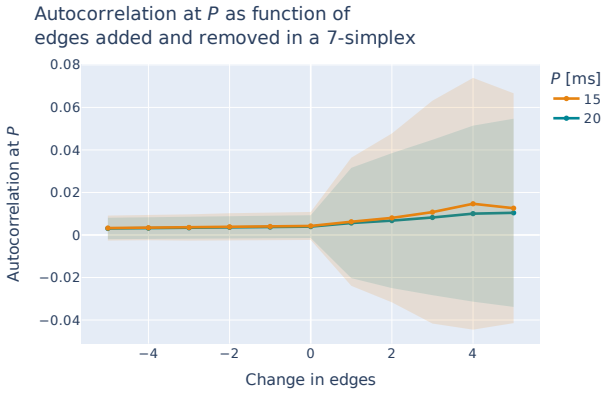


(b) Standard deviations shown in shaded regions.

Figure 18: Average autocorrelation of sink neuron at P, for $P = 15$ and 20 ms, as function of simplex size. Both figures show the same data, with the standard errors and the standard deviations shown in shaded regions in subfigs. (a) and (b) respectively.



(a) Standard errors shown in shaded regions.



(b) Standard deviations shown in shaded regions.

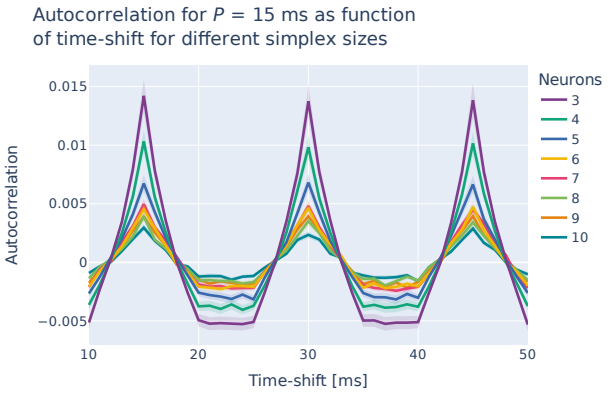
Figure 19: Average autocorrelation of sink neurons with P for $P = 15$ and 20 ms, as function of the number of random edges added or removed in 7-simplices. Both figures show the same data, with the standard errors and the standard deviations shown in shaded regions in subfigs. (a) and (b) respectively.

in the autocorrelation for the larger simplices. This should not be too surprising, for in accordance with the previous findings, the causal link between source and sink neurons are strongest in the smaller simplices as a result of the lack of other input to the sink neurons. Thus, it is to be expected that when considering time-shifts of P, the spike patterns in the source neurons as function of the activity in the sink neurons are also strongest for the smaller simplices. The cause of the high variance might be that given the lack of other input, the effect of differences in the synaptic strengths will be greater.

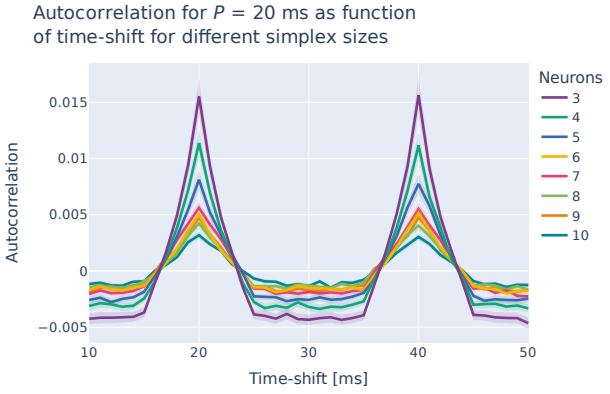
The same analysis was then performed with respect to the number of edges added or removed in complete 7-simplices, and the resulting plots shown in fig. 19. Somewhat surprisingly, removing edges had little effect on the average autocorrelation, whereas adding edges led to a very significant increase. Further, the standard deviation also exploded with the increase in the number of edges. The fact that removing edges has little effect suggests that the main source of the autocorrelation comes from the direct link from source to sink, which is reasonable given that we are only considering the autocorrelation at P. The question is then what causes the strong increase in the autocorrelation with the added edges. Adding extra edges has the effect of introducing causal loops in the network, so it is not too surprising that this leads to additional repeating patterns in the spike trains. Further, these additional edges can be both excitatory and inhibitory, and the increase in variance suggests that the change in autocorrelation is dependent on the type and weights of the additional edges.

Next, the autocorrelation was plotted as a function of time-shift for different simplex sizes, as shown in fig. 20. The standard errors are included in the top row, the standard deviations in the bottom. The maximum autocorrelation is, as expected, at integer multiples of P for all sizes, and the strength is directly proportional with the size of the simplex as found before. More interestingly here is what happens between the peaks, for the smaller simplices show a corresponding extremity in the opposite direction, where they are the most strongly anticorrelated between peaks. The bigger the simplices, the more flattened and centred around zero is the autocorrelation of the sink neurons. To understand this, it is useful to remind oneself of how to interpret correlations, which are in the range [-1, 1]. Importantly, negative values are equally indicative of repeating patterns as the positive values, they simply show that the variables take opposite values. Thus, what fig. 20 shows is that it is the smaller simplices which show the strongest autocorrelation in their time-series across all time-shifts, refuting the hypothesis that the bigger simplices might show longer sequences of repeating patterns. Looking at the graphs in the bottom row, which includes the standard deviations, it is also clear that there is significant variation in the autocorrelation within each group, and this is larger for the smaller simplices.

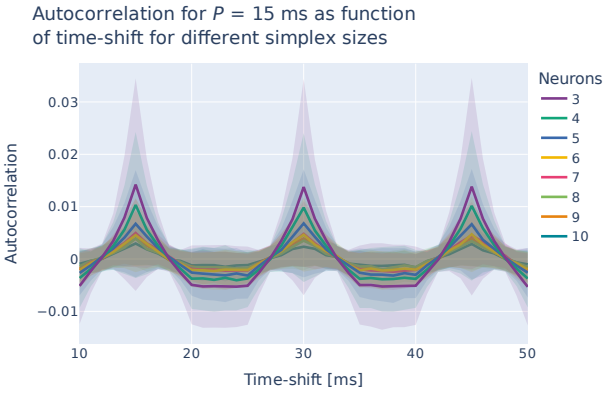
Finally, the same analysis was performed on the spike-trains of sink neurons in 7-simplices where edges had been added or removed. Given the previous findings, which suggested a much stronger effect of adding edges than removing them, the two cases were treated separately. Only the stimulation with $P = 15$ ms was here considered, since no substantial difference was found between the two cases earlier. The resulting plots are shown in fig. 21, with standard errors included in the top row and standard deviations in the bottom. We see that the autocorrelation as function of time has almost exactly the same shape independently of the number of edges added, but its value increases with increasing number of edges. This is thus a different type of dependency from that seen when the size of the simplex was the



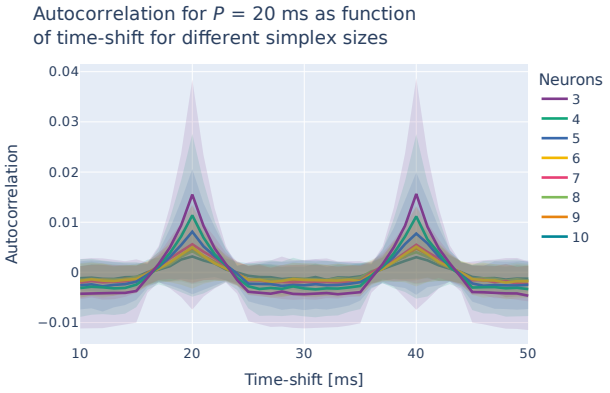
(a) Stimulation of source neuron every 15 ms, standard errors in shaded regions.



(b) Stimulation of source neuron every 20 ms, standard errors in shaded regions.

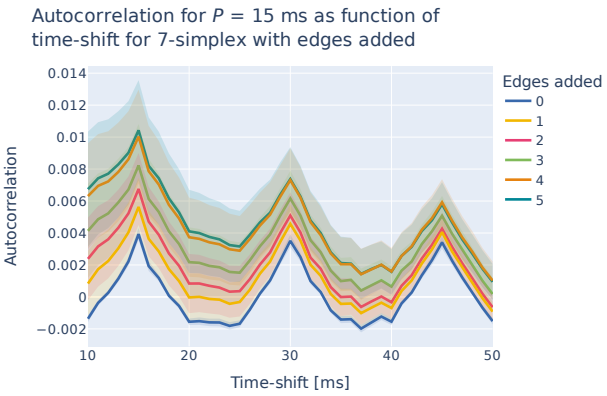


(c) Stimulation of source neuron every 15 ms, standard deviations in shaded regions.

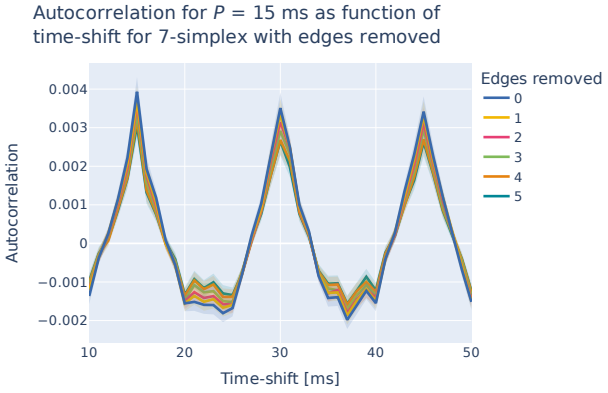


(d) Stimulation of source neuron every 20 ms, standard deviations in shaded regions.

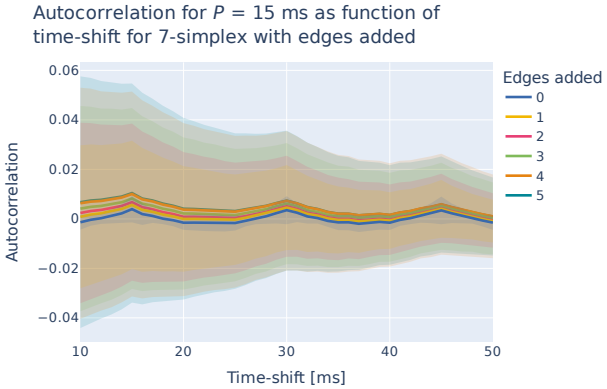
Figure 20: Average autocorrelation of sink neuron with $P = 15$ and 20 ms as function of time-shift for simplices of different sizes. Standard errors included in the shaded regions in subfigs. (a) and (b) and standard deviations in subfigs. (c) and (d).



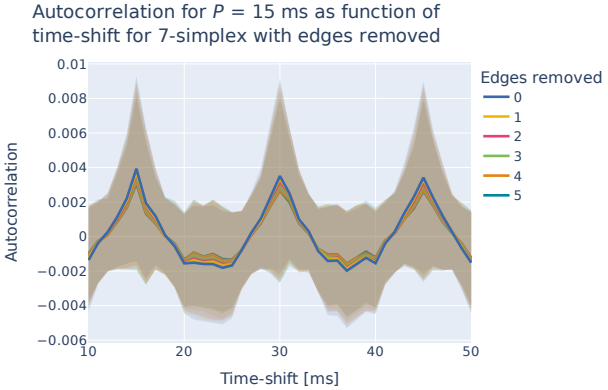
(a) Edges added, standard errors in shaded regions.



(b) Edges removed, standard errors in shaded regions.



(c) Edges added, standard deviations in shaded regions.



(d) Edges removed, standard deviations in shaded regions.

Figure 21: Average autocorrelation of sink neuron with $P = 15$ ms as function of time-shift for a 7-simplex with different number of edges added (left) and removed (right). Standard errors included in the shaded regions in subfigs. (a) and (b) and standard deviations in subfigs. (c) and (d).

variable, where the autocorrelation flattened out for bigger simplices. Whereas increasing the size of the simplex *reduces* the autocorrelation and pushes it toward 0 for all time-shifts, adding additional edges (causal loops) had little effect on the shape of the curve, but increases its value. However, looking at fig. 21c, where the standard deviation is included in the shaded regions, it is again clear that the within-group variance of the networks with the same number of edges added is substantial, making it dubious whether it is really meaningful to treat these as unified functional groups with respect to the autocorrelation of their sink neurons. Interestingly, the effect of *removing* edges, as shown in fig. 21b, is more or less negligible, and the autocorrelation curves mostly overlap. This pattern is still apparent when the standard deviations are included, as in fig. 21d, such that there is hardly any discernible difference between the distributions, suggesting that they all belong to the same functional group with respect to their autocorrelation.

Analysis of Variance Given that the Mann-Whitney U test had been used throughout thus far, this was used here as well without further tests of normality and equality of variance, thus making the results more comparable with the previous analysis. Given that the patterns in the findings were very similar for both of the stimulation rates investigated, only the datasets with $P = 15$ ms were considered in this analysis.

It was first tested whether the difference in simplex size was significant for time-shifts of P and $1.5P$, i.e. at the two extremities. The results showed that only the simplices consisting of 3 and 4 neurons were significantly different from the other sizes, for both time-shifts.

When looking at the 7-simplices with different number of edges added and removed, all the simplices with edges removed were grouped together and all those with edges added were grouped together. These groups were then compared with each other, as well as with the complete simplices. None of the groupings were significantly different from the others, with the exception of the group of networks with edges added compared with the group of networks with edges removed, when considering a time-shift of P. It should be noted that the extremely high variance in the datasets with edges added compared to that in the other datasets makes it difficult to obtain statistically significant results in this case, so little can be inferred from this.

4.6.2 Discussion

The main hypothesis tested in this section was that the autocorrelation of the activity in the sink neurons in networks with periodic input to the source neurons increases with simplex size and completeness of the simplex. Neither could be proven, and the autocorrelation was rather found to be highest for the smallest simplices across all time-steps. However, one should be cautious with concluding too much from this finding, since this is one instance where the network size and in-degree of the sink neurons can plausibly be assumed to influence the results. In the smaller isolated simplices, the sink neurons only receive input from a limited number of other neurons. As the in-degree of the sink neurons increases, it is reasonable to assume that this introduces a higher degree of randomness in their activity, since the firing rates in their pre-synaptic neurons are unlikely to be synchronised. This introduces random noise in the activity of the sink neurons, thereby reducing the autocorrelation and the signal-to-noise ratio. In more integrated networks, the in-degree can be higher also

for smaller simplices, and it is unknown whether the relationship between simplex size and autocorrelation would remain the same in such cases. The results here obtained should thus be seen as a limiting case where the noise is kept at a minimum for simplices of all sizes, so likely reflect the upper limits of the autocorrelations for the simplices.

Regarding the completeness of the simplices, the main finding was that removing edges had very little effect on the autocorrelation, whereas adding edges increased it (although the variance here was too high to make this finding statistically significant). This is an interesting finding given how the directed simplices are defined. All the 7-simplices with added edges, which show a significant difference in the autocorrelation, are treated as the same structure according to the definition of directed simplices, whereas the different graphs with edges removed, which show almost the exact same functionality when it comes to autocorrelation, are classified as different structures. It thus seems safe to conclude that autocorrelation is not the functional trait which justifies the structural classification and pursuitworthiness of directed simplices.

4.7 Summary and Conclusion

Thus far, no positive results in favour of a special function of simplicial structures in biological neural networks has been found. We were unable to show that higher-order simplices have a more important functional role according to the metrics tested. Rather, for all three metrics, namely relative average treatment effect, transfer entropy and autocorrelation, the smaller simplices scored higher throughout. Further, there was little evidence to support the strict structural definition of the directed simplices with respect to these metrics. Rather, a continuous relationship between the completeness of the simplices (in terms of edges added or removed) and the different functionalities were found both for ATE and transfer entropy, whereas removing edges hardly had any effect on the autocorrelation.

The main finding was probably the high levels of variance in all the datasets with respect to the different metrics. This shows that grouping the structures according to their simplex size is likely not the most informative way of functionally categorising these structures. This argument was further supported by the analyses of variance, where hardly any of the groupings considered were statistically significant.

However, from this we are only justified in rejecting the specific hypotheses about function which were tested. As discussed, *function* is a vague term, so these results are by no means sufficient to reject the looser hypothesis that these structures entail *some* special function. It might simply be the case that the specific function which they entail has not been tested for here.

5 Machine Learning Model

5.1 Motivation

A more general and model-independent way of thinking about the function of the simplicial structures is to say that their having a function implies that there exists some (potentially deterministic) mapping from structure to function, $\mathcal{S} \rightarrow \mathcal{F}$. This would imply a one-to-one or a many-to-one mapping from structure to function, meaning that the same structure always gives rise to the same function, but that multiple structures may give rise to the same function.

Similarly, one could look for an inverse mapping, $\mathcal{F} \rightarrow \mathcal{S}$, where one assumes that each function corresponds to a single structure, but that different functions could arise from the same structure. This is a less intuitive interpretation, since one would normally consider the structure as being the cause of the function, rather than the other way around. Assuming determinism, it would thus be difficult to imagine a scenario where the same structure could give rise to different functions, so for this to be plausible, one would presumably have to include some other variables $\mathcal{U}$ in the mapping, such that we have $\mathcal{F} \rightarrow \mathcal{S} + \mathcal{U}$. In our case, $\mathcal{U}$ could represent the synaptic weights or random input noise, to mention just a few. Thus, one could understand function in either of the two ways.

The crucial point here is that the function $\mathcal{F}$ is left undefined, so it is not possible to construct a general test for the hypothesis that it exists. One can only come up with a range of different candidates and test these, as was done in the section above, with negative results. However, this is one case where the use of machine learning could be valuable. In the not so recent past, machine learning models have successfully been used to find mappings from one set of variables to another, even in cases where humans themselves are unable to find this functional mapping. A prominent example of this is the relatively recent (2020) machine learning model AlphaFold2, which is able to predict the 3D structure of folded proteins from their linear amino-acid sequence [16], something which humans have thus far been unable to do. Implicit in the discovery of such a model is the fact that there must exists some (at least partly) deterministic mapping from the amino-acid sequence to the folded protein. In other words, the information about the 3D structure of the folded protein is *somehow* encoded in the amino-acid sequence, it is simply us who are ignorant of it.

For our purposes, if it turned out to be possible to train a machine learning model to successfully perform the mapping between function and simplicial structure (in either direction), this would be a strong result in favour of the interdependency between simplicial structure and function of such networks. As a result of the high levels of stochasticity in the simulated activity, it is rather unrealistic to be able to train a model to accurately perform the mapping from structure to function. Further, on more pragmatic grounds, if the intention is that this could also provide a tool useful in the analysis of experimental data, one must also take into consideration what observational data is usually available and what one wants to predict. As discussed in section 2.1.2, spike-trains of individual neurons are obtainable, whereas complete structural or effective connectomes with synaptic weights are near impossible to measure experimentally in vivo. Thus, in this case, we were more interested in the possibility of finding a model which was able to perform the mapping $\mathcal{F} \rightarrow \mathcal{S} + \mathcal{U}$, as well as differentiating between $\mathcal{S}$ and $\mathcal{U}$. This was the overarching purpose

of the investigations and analyses presented in the current section.

5.2 Edge Regressor

The code for the models employed is available at [27]. Using the message passing algorithm in GNNs, its functionality is to perform the mapping from the time series $\mathbf{X} \in \mathbb{R}^{N\times T}$ to a prediction of the connectivity matrix $\hat{\mathbf{W}}_\mathbf{0} \in \mathbb{R}^{N\times N}$. The i$^{\text{th}}$ row of the time-series matrix $\mathbf{X}$ is denoted x_i, and represents the complete time-series of node i. The rows of the connectivity matrix can be taken as the node embeddings $\mathbf{w}_i$, where each element of the vector represents the edge weight of the outgoing edge to the receiving node in the network. Thus, $\hat{\mathbf{W}}_\mathbf{0} = [\hat{\mathbf{w}}_1, \hat{\mathbf{w}}_2, ..., \hat{\mathbf{w}}_i, ..., \hat{\mathbf{w}}_N]$. We denote the edge weight from node i to node j $(W_0)_{i,j}$. Further, the adjacency matrix is unknown, so all elements in $\mathbf{A}$, with the exception of the dioagonal, were set to 1 prior to training, such that messages were passed in both directions between all the neurons in the network. Using the formalism from section 2.4, where $\mathcal{N}(i)$ denotes the set of all the neighbours of node i, this means that $\forall j(j \neq i \rightarrow j \in \mathcal{N}(i))$. The model found to be capable of performing the mapping $\mathbf{X} \rightarrow \mathbf{W_0}$ can be expressed as

$$\hat{\mathbf{w}}_i = MLP_2 \left(\Big\|_{j\in\mathcal{N}(i)} MLP_1 \left(\frac{100}{T} \Big\|_{t=1}^{C} x_i \cdot P^t x_j \right) \right). \tag{7}$$

T is the total number of time-steps, C is the number of time-shifts considered in P. This is the hypothesised coupling window c_w, which was here set to 10 ms, since this was known. The aggregation function used was concatenation.

The expression inside the first MLP requires further elaboration. The aim was to learn the causal effect between the neurons from their spike trains. This was done by multiplying their respective time series for up to C time-shifts. The complete time series of the hypothesised sender node i, i.e. x_i is multiplied (dot product) with the time series of the hypothesised receiver node, x_j, for different time-shifts applied to the receiver node. The time-shift is represented by the shift-operator P^t, where t denotes the time-shift. Thus, for time-shift 3, the resulting number from the dot product tells us something about the rate of cofiring between the node i and node j three time-steps after node i fired. The resulting dot products for up to C time-steps are then concatenated into a single vector q_j, which is what is passed through the first MLP. Thus, the first MLP learns the mapping from the vector q_j to a single number, which should somehow represent the causal effect of node i on node j. For each node i, this is done for every other node in the network and the results concatenated into another vector of length N. This vector is then passed through the second MLP, allowing the final prediction of $\hat{\mathbf{w}}_i$ to take potential interdependencies between the different q_js into account.

The cost function used was mean squared error (MSE), such that

$$C = \frac{1}{N^2} \sum_{i=1,j=1}^{N} ((W_0)_{i,j} - (\hat{W}_0)_{i,j})^2 \tag{8}$$

In previous analyses, the model has been found to perform well at the task of predicting the connectivity matrix of networks of different types and sizes. As an example of the predictions provided by the model, a sample of a prediction from the test set of a small-world network containing 30 neurons is shown in fig. 22.

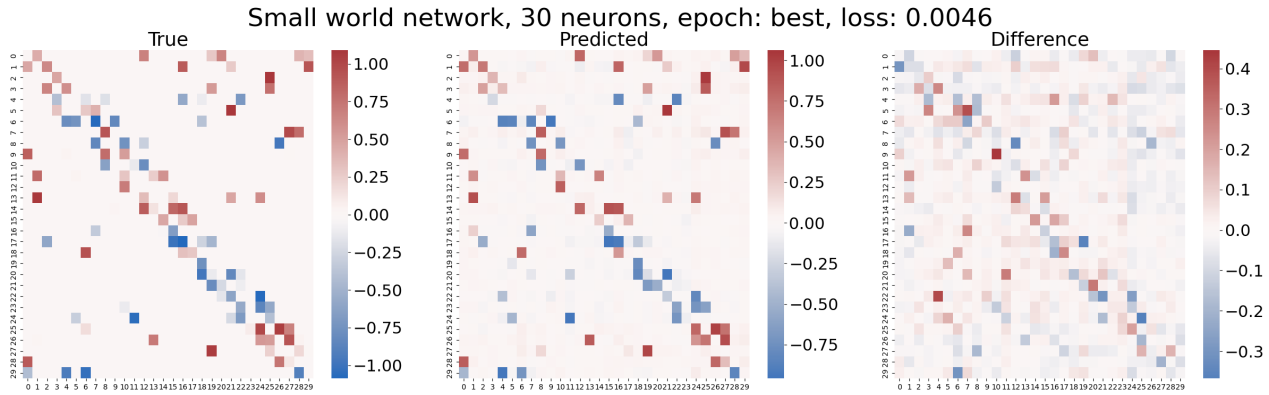


Figure 22: Sample from the test set of a model trained on small-world networks of 30 neurons. The figure shows the true connectivity matrix, along with the prediction and the difference between the two.

5.3 Hyper-parameter tuning

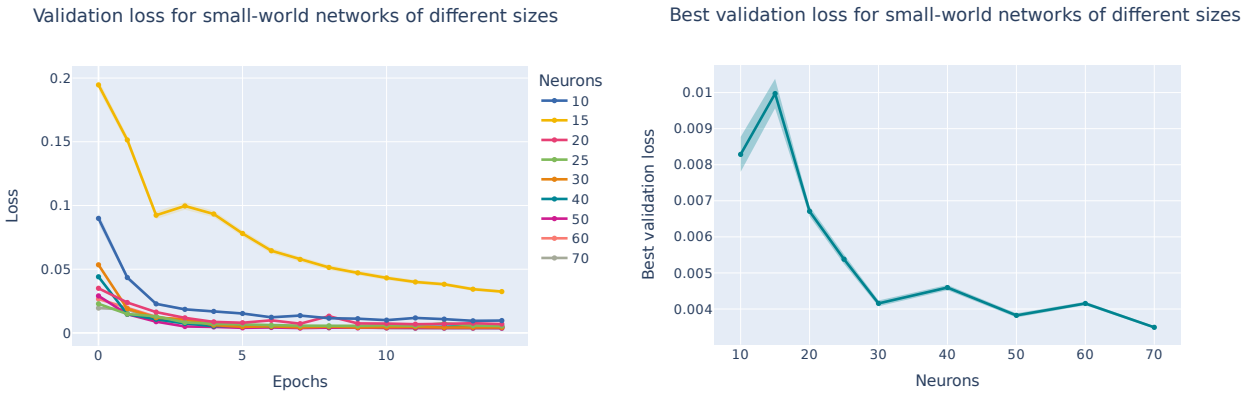
The only hyper-parameter in the model is the learning rate. Some rudimentary testing was done for a few different network sizes, and the outcomes were fairly constant and independent of small changes in the learning rate. Thus, this was kept constant at 0.003. This lack of rigorous testing was justified by the fact that the purpose of the investigations was primarily to compare different results obtained from the same algorithm, rather than to find the best possible model for predictive purposes. The analysis consisted in comparing different models trained using the same algorithm, so as long as the hyper-parameters were kept constant across models, their actual values were of lesser significance. The maximum number of epochs was set to 50, as most of the models converged far before this, after about 20 epochs. The batch size was set to 1, mainly because of limited GPU memory.

5.4 Model Reliability

Before testing whether the model was able to reconstruct the simplicial structures from the activity, we performed some different tests on the reliability of the model when set to predict the connectivity matrix. This was to better be able to analyse the later results, and to rule out other sources of error.

5.4.1 Scaling

To test how well the ML algorithm scales with network size, models were trained for both small-world and random networks with sizes ranging from 10 to 70 neurons. Bigger networks



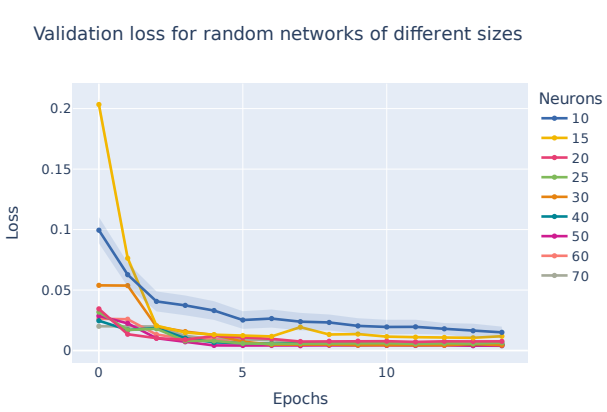
(a) Loss as function of training epochs for different network sizes.

(b) Best validation loss as function of network size.

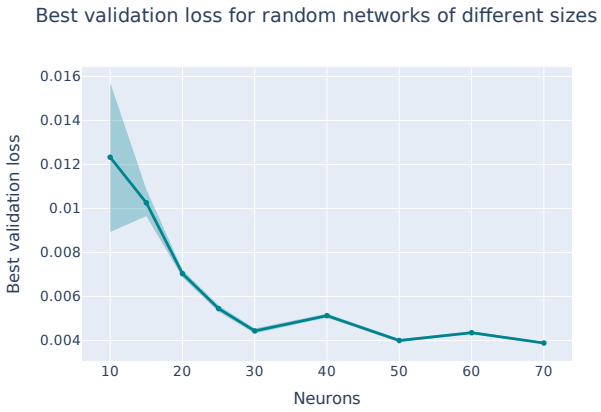
Figure 23: Validation loss in small-world networks of different sizes. Standard errors shown in shaded regions.

were not tested due to computational limitations. In fig. 23, the losses for different network sizes of small-world networks are shown. In fig. 23a, the validation losses are shown as function of epochs for a range of different network sizes. As can be seen, most of the networks converge towards a similar loss, regardless of size. fig. 23b supports this further. Here, the minimum validation loss across epochs is plotted as function of network size, and as can be seen, the model scales well. Why the loss is significantly higher for the network containing 15 neurons is unknown, but is likely a random fluctuation in the datasets. The same pattern is found for the random networks, as shown in fig. 24. Indeed, the minimum validation loss actually decreases quite substantially with the simplex size.

This is slightly peculiar, and is likely related to the fact that the number of edges in the networks do not scale linearly with the number of neurons. This introduces an increasing bias towards edges being set to 0, which is a problem in this comparison. To investigate whether this was the case, plots of the predictions for the random networks containing 30 and 70 neurons are shown in fig. 25 and fig. 26. Looking at these, one can see that although the MSE loss is higher for the smaller network, the maximum difference relative to the true values is higher for the network containing 70 neurons. This can be seen from the scales of the colour-maps. Whereas the true values of the network containing 30 neurons are in the range $[-1.0, 1.0]$ and the errors in the range $[-0.4, 0.4]$, the true values of the network containing 70 neurons are in the range $[-0.6, 0.6]$, yet the errors are in the same range as for the smaller network. This suggests that the individual predicted values are less reliable in the bigger networks, such that the average MSE loss is somewhat misleading, and that the model does not in fact scale as well as the previous results would suggest. This is not particularly surprising, since the increase in size increases the complexity of the problem, but is certainly something worth being aware of.



(a) Loss as function of training epochs for different network sizes.



(b) Best validation loss as function of network size.

Figure 24: Validation loss in random networks of different sizes. Standard errors shown in shaded regions.

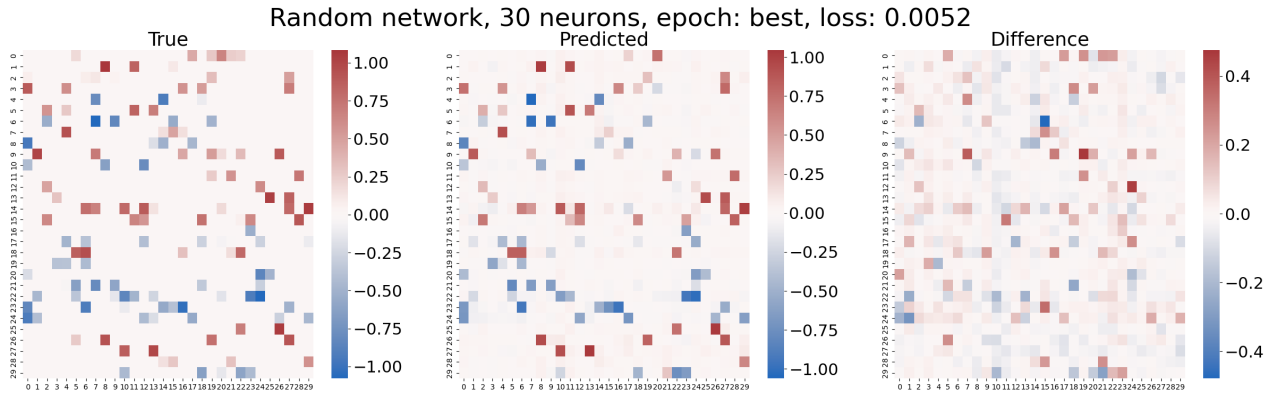


Figure 25: Sample from the test set of a model trained on random networks of 30 neurons. The figure shows the true connectivity matrix, along with the prediction and the difference between the two.

5.4.2 Confounding Variables

As discussed in section 2.3, a major challenge in causal inference is to distinguish between true causal relations and correlations which arise from random fluctuations or confounding variables. As mentioned, various methods have been introduced for causal inference, and which one to use depends on the problem. The hope is here that the outer MLP is able to, without further supervision, learn an appropriate method for causal inference also in the presence of confounding variables.

When training the models on the complete graphs, there is random noise in the data, but no confounding variables. Good predictions thus show that the model is able to learn to distinguish between causal relations and noise, but this is normally not sufficient for establishing causality. To test whether the model is also able to distinguish between actual causal links and confounding variables, we introduced confounding variables in the datasets.

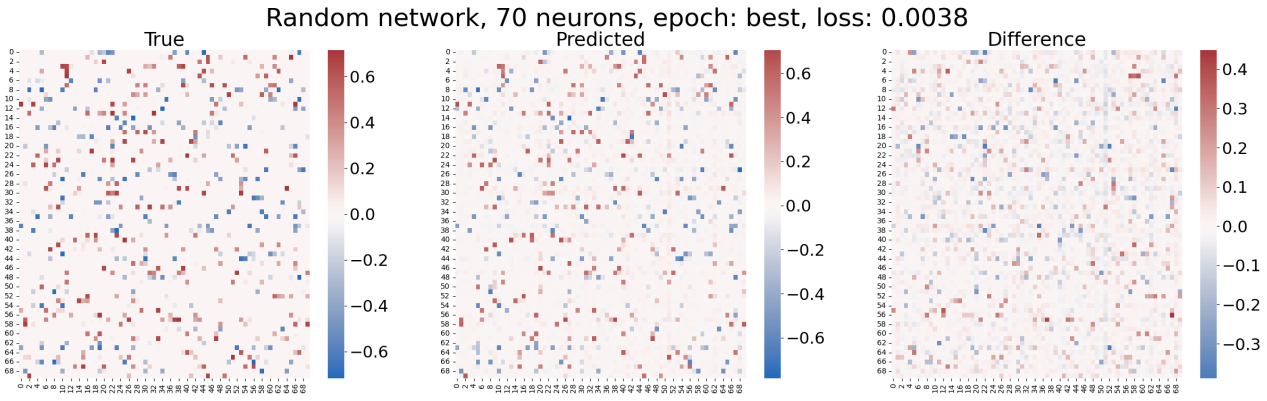


Figure 26: Sample from the test set of a model trained on random networks of 70 neurons. The figure shows the true connectivity matrix, along with the prediction and the difference between the two.

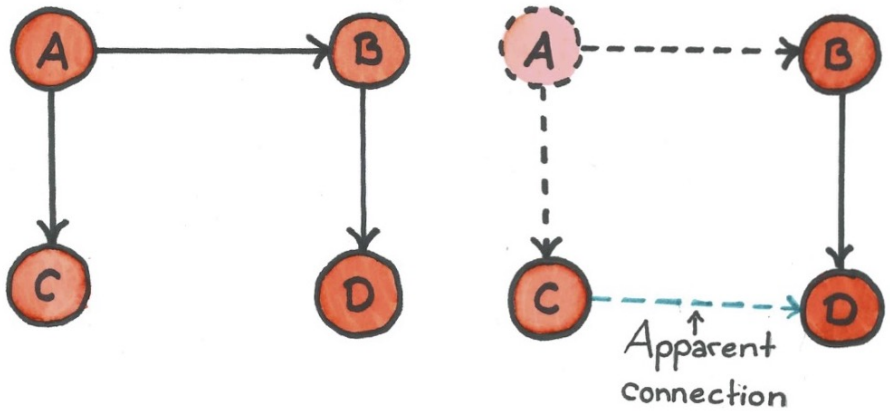
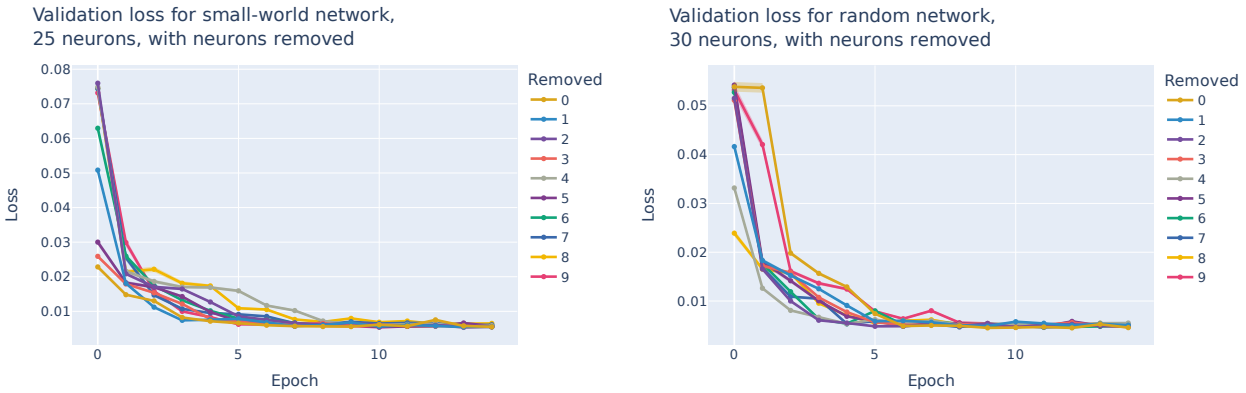


Figure 27: An example of a network before and after a neuron has been removed from the available data. The network to the left is the complete network containing four neurons, and is the one used to generate the spike-trains. All information pertaining to neuron A is then removed from the connectivity matrix and the spike-trains in the data-sets. If we assume all synapses are excitatory, the structure of the true network means that neuron A then becomes a confounding variable, and without the information about its spiking-pattern available, if one only looks at the correlations, it might easily look like there is a causal connection from node C to node D. The aim of the current test was thus to see whether the model was able to predict only the true connection from B to D.

This was done by removing datapoints in the data, both in the spike-trains and the connectivity matrices. Specifically, m randomly selected neurons were completely removed from the datasets. This was done by removing these neurons' spike trains from the $\mathbf{X}$-matrix and removing all of their row and columns from the connectivity matrix. Thus, if the original network contained n neurons, the original $\mathbf{X}$ had dimensions $n \times T$ and the original $\hat{\mathbf{W}}_\mathbf{0}$ $n \times n$. In the adjusted dataset, the $\mathbf{X}$ is of dimension $(n-m) \times T$ and the $\hat{\mathbf{W}}_\mathbf{0}$ of dimension $(n-m) \times (n-m)$. Note that the removed neurons were not removed from the simulation,



(a) Small-world network of 25 neurons with up to 9 neurons removed

(b) Random network of 30 neurons with up to 9 neurons removed

Figure 28: The validation loss plotted as function of epoch for networks with different numbers of neurons removed. Standard errors shown in shaded regions.

however, so their causal influence was still present in the dataset. However, since they were unknown to the model, they worked as unobserved confounding variables. An example of such a removal is illustrated in fig. 27. For each combination of n and m, a new model was trained to see whether the training algorithm would make it possible for the model to learn distinctly causal, and not just correlational, relationships.

This analysis was done for small-world networks of 25 neurons and random networks of 30 neurons. In both cases, up to 9 neurons were removed. The resulting validation losses as function of epoch for different numbers of neurons removed are shown in fig. 28. As can be seen, the loss converges to about the same values for all the networks, showing that the model does indeed pick out the distinctly causal relations between the neurons, so is able to perform the causal inference in contexts where both noise and confounding variables are present.

5.5 Simplicial Thresholding

To test whether the models were able to pick out the higher order simplicial structures, a pre-processing scheme of simplicial thresholding was introduced. Here, nothing was done to the spike-trains, but the connectivity matrices were adjusted based on the maximal simplicial complex of the network. For each threshold, only edges which took part in simplices of a certain dimension or above were kept, the others were set to zero. Since almost all edges are at least part of a 2-simplex, the lowest threshold tested was that of 3-simplices. For example, when using a threshold of, say, 5, only the edges which form part of simplices containing at least 5 neurons are kept. All other edges are set to 0. An illustration of filtering of 3-simplices is shown in fig. 29.

Two different approaches were taken to see whether the models were able to learn the simplicial structures of the networks. First, the simplicial thresholding was applied to test datasets and sent through models trained to predict the complete connectivity matrix. Second, the thresholding was applied to the training datasets as well, and it was tested whether

it was possible to train models to only learn the thresholded connectivity matrices.

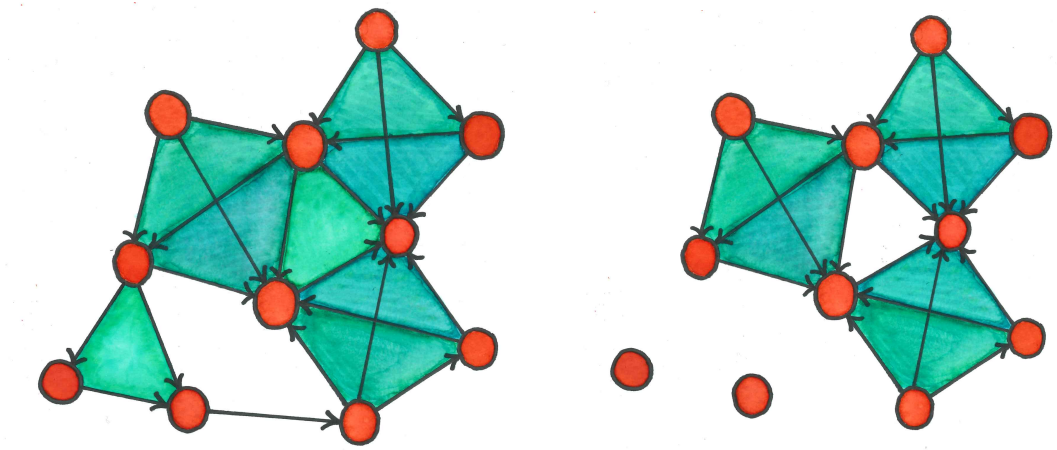


Figure 29: An example of simplicial thresholding of 3-simplices (4 neurons). The figure to the left shows the original network, whereas the figure to the right shows the network with the remaining edges after simplicial thresholding of 3-simplices has been applied. Note that the 2-simplex in the middle nonetheless remains, as all of its constituent edges form part of the surrounding 3-simplices.

5.5.1 Testing

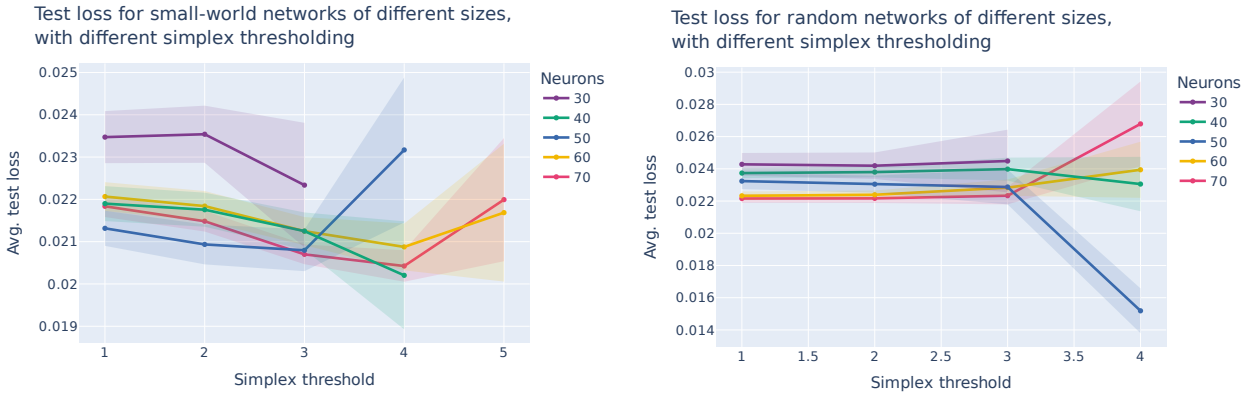
First the intention was to see whether the models trained on the complete simplices showed any difference in accuracy in predicting the edge weights of the different edges based on their simplicial role. The idea was to see whether the models were more accurate in their predictions of the edge weights of the edges which form part of higher-order simplices than those which are only part of smaller simplices. If this were to be the case, it would be an indication that whatever information the model uses to predict the edge weights is somehow more prevalent or easily extractable when the edges form part of higher order simplices.

The original models obtained when training on the complete connectivity matrices were used, and the test-set was pre-processed and used in the analysis. For each pair of prediction $\hat{\mathbf{W}}_\mathbf{0}$ and ground truth $\mathbf{W_0}$, the pair was sent through the simplex filtering process such that both were filtered based on the maximal simplicial complex of $\mathbf{W_0}$. Denote this thresholded set $\hat{\mathbf{W}}_{\mathbf{0,t}}$ and $\mathbf{W_{0,t}}$ for threshold t. This was done for each filtering-threshold up to a certain value. Consequently, they both have the non-zero values at the same indices, though the actual weights predicted might differ. When calculating the loss, a slight alteration of the MSE loss was used. Instead of dividing by the total number of potential edges, $N \times N$, the sum of squares was divided by the number of non-zeros edges in the thresholded $\mathbf{W_{0,t}}$, call this $E_{NZ,t}$. The loss at threshold t was thus:

$$C_t = \frac{1}{E_{NZ,t}} \sum_{i=0}^{N} \sum_{j=0}^{N} ((\hat{W}_{0,t})_{i,j} - (W_{0,t})_{i,j})^2 \tag{9}$$

Note that for higher thresholds, some of the losses became 0 since there were no simplices remaining. These losses were excluded when calculating the averages, since only the losses on the non-zero edges were of interest. Importantly, a filtering of 1 makes no difference to the $\mathbf{W_0}$ matrix, since every edge is per definition part of a 1-simplex. However, it *does* change the prediction $\hat{\mathbf{W}}_{\mathbf{0,t}}$, since all the edges which are incorrectly predicted to be non-zero are set to 0 in the filtering process. The loss should therefore decrease significantly even at the first step as a result of this, whereas what we are really interested in is the dependency after that. Thus, the 1-threshold was included as the point of reference for how accurate the model was in predicting the non-zero edges of the complete connectivity matrix.

This was tested for small-world and random networks of different sizes, and the resulting plots of test loss as function of threshold is shown in fig. 30. As can be seen, no general pattern of dependency of loss on simplex threshold was found, and the changes detected were marginal relative to the baseline. This shows that how well the model learns a certain edge weight is independent of that edge's simplicial role. The hypothesis that the models are better at predicting edges which form part of higher order simplices is thus falsified, though little more can be said about the functionality of the edges themselves from this negative result. Further, given that these models were trained to learn the complete connectivity matrix with equal importance of all the edges, this result is not particularly surprising.



(a) Small-world networks

(b) Random networks

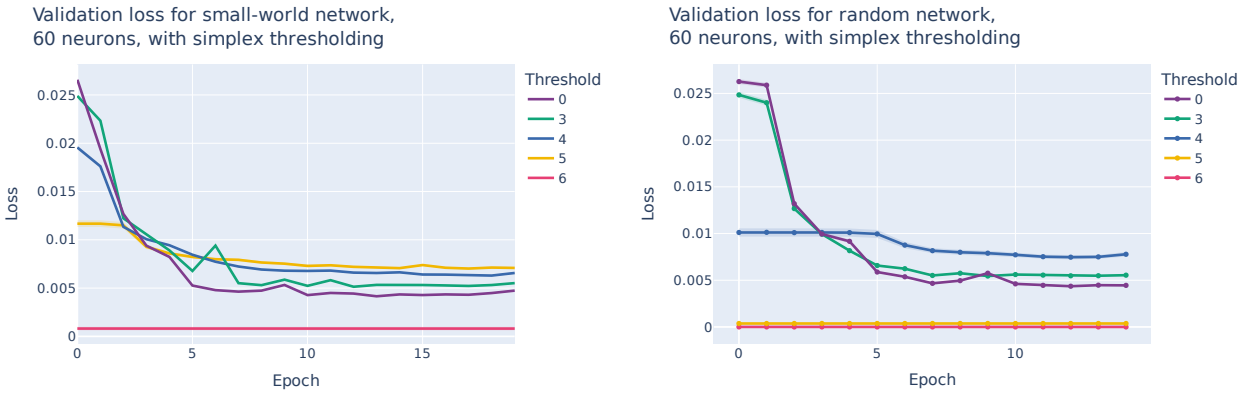
Figure 30: Average test loss on non-zero values as function of simplex thresholding for small-world and random networks of different sizes. All the models were trained to predict the complete connectivity matrices. Standard errors shown in shaded regions.

5.5.2 Training

To better test whether the edges in higher-order simplices are functionally distinguishable from those in smaller simplices, we attempted to train new models to only predict the filtered connectivity matrices. This was done by performing the simplex thresholding to the entire dataset prior to the training. To test whether it was possible to train the models to only pick out simplicial complexes of a certain dimensionality, a new model was trained for datasets with each threshold. A positive result would then imply that these higher-order structures are somehow detectable in the time-series, and thus that they must have some functional

coherence which the model is able to learn, even though we have been unable to identify this. Importantly, a negative result is in this case not sufficient to falsify the hypothesis that they *do* have some functional role, since all this tells us that *this specific* machine learning algorithm is unable to learn to pick out these structures, which does by no means imply that the information is absent from the input data.

This analysis was done for networks containing 60 and 70 neurons, both small-world and random. Thresholding up to 5-simplices were tested. Note that the standard MSE loss was here used, since we are also interested in the models' ability to correctly predict the absent edges. The resulting plots of validation loss as function of epochs for different simplex thresholds are shown in fig. 31. The most obvious result is probably that the highest filtering gives the lowest loss. This is simply because there are hardly any edges left in these networks, such that the models primarily learn to correctly predict zeros. For the networks where there are still edges present, however, one can see that the models perform slightly worse with increasing thresholding. This suggests that the models are unable to pick out only the higher-order structures. Further, this difference is likely even more significant when we take into account the increasing number of zeros in the matrices where thresholding has been applied. Previously, we found that the loss decreased with the percentage-wise increase in zero-weights. This is likely the case here as well, so this is a clearly negative result with regard to the hypothesis that the models will be able to learn to pick out higher order functional relations from the time-series.



(a) Small-world network of 60 neurons

(b) Random network of 60 neurons

Figure 31: The validation loss plotted as function of training epochs for small-world and random networks, both containing 60 neurons, with different simplex thresholding applied. Standard errors shown in shaded regions.

Before rejecting the hypothesis, a final possibility was considered. It could be the case that in order for the function of the higher order simplices to become observable, one must consider greater time-shifts than the coupling window of 10 ms. In the above analysis, looking at eq. (7), the maximum shift of the vector P was 10 time-steps (i.e. $T = 10$ ms). This was a fairly arbitrary choice, motivated by the fact that coupling window c_w from the GLM was known to be 10 ms. Thus, the maximal time-shift was increased to 30 ms, and the above analysis repeated. The resulting plot for small-world networks of 70 neurons is shown in fig. 32. As can be seen, the exact same pattern is found as with the time-shift of 10 ms,

where the loss increases with the thresholding, showing that the model remains unable to differentiate between the edges based on their simplicial role. The same was found for both small-world and random networks of other sizes.

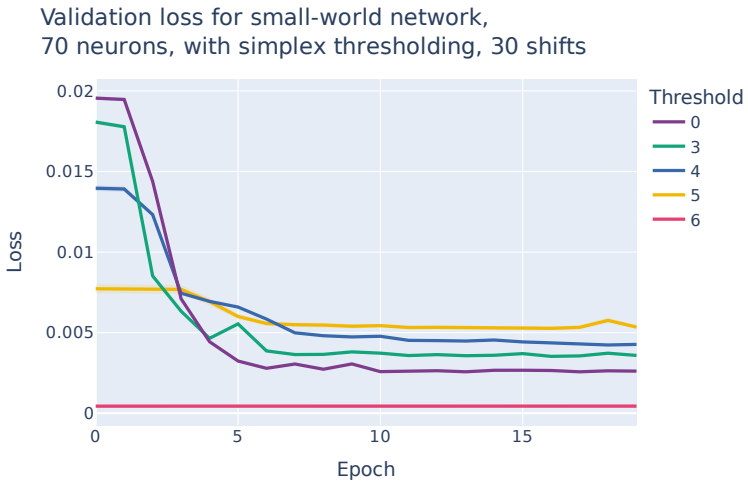


Figure 32: The validation loss plotted as function of training epochs for small-world network of 70 neurons when using a maximal time-shift in P of 30 ms, with different simplex thresholding applied. Standard errors shown in shaded regions.

5.5.3 Discussion

So what can be said about the function of the simplices based on these results? In reality, very little can be inferred from a null result in an investigation like this. The only thing which has been shown is that this specific machine learning algorithm, which has been shown to be able to find the dyadic effective and structural connections between neurons, is unable to differentiate between the connections which form part of higher order directed simplices and those which do not. It is, however, an indication that the hypothesised function which arise from the simplicial structures is not instantiated in the relationships between pairs of neurons. In other words, given only the spike-trains of two neurons, it might be impossible to infer their simplicial relation. Instead, the function might be instantiated at a higher level, as a property which might only become evident when looking at certain relationships between the spike-trains of many neurons simultaneously.

This is related to a more general problem in the application of machine learning models, especially when they are being used for explanatory purposes such as this. A commonly raised criticism of machine learning models is their double opacity, where both what is learned and how it is learned remains opaque to the programmer [7]. This lack of knowledge of what the model actually learns and bases its predictions upon makes it near impossible to draw any inferences about the dataset just from the model's inability to learn certain features. In our case, it is unclear to what extent the current algorithm is able to pick up on higher-order relationships, and testing for this is very difficult. The only indication we have that it picks up on some more integrated relationships than the purely dyadic conditional probabilities of the firing rates is that it is able to differentiate between correlations resulting

from confounding variables and direct causal links. To do so, it seems reasonable to assume that some analysis of higher-order relationships in the network is necessary. At the very least, it must be able to consider a neuron's relationship to multiple different neurons at the same time, and how these relate to each other. Hence, if the information about the functional properties of the simplicial structures is available at this level, the model should have been able to learn it. Nevertheless, this is too vague and uncertain an argument to conclusively say anything of substance about whether the simplicial information is present in higher-order functional properties of the networks.

6 Conclusion and Outlook

6.1 Conclusion

In this project, we investigated whether directed simplices in biological neural networks are associated with certain functions. This was done in two parts, first using methods and ideas from the field of causal inference to test for three specific functional properties, and second, with a more model-independent perspective, using a machine learning model. This constituted a novel approach to the study of the functional role of such structures in biological neural networks. In previous work, such as [47] and [21], only correlations in the functional connectivity in such networks have been considered. In contrast, the main emphasis in this project has been on analysing the causal effects of activity in source neurons in higher-order simplices on the dynamics of the other neurons, primarily that of the sink neurons in the same simplices.

In sum, no positive results were obtained regarding the increasing functional significance of higher-order simplices in the simulated activity in the networks studied. Consequently, we also have no indication that they can be treated as (semi-)independent causal units in the networks. This held both when considering the specific metrics of relative ATE, transfer entropy and autocorrelation, as well as for the model-independent approach taken using the machine learning model. Unfortunately, however, very little can be ascertained from the negative results obtained in this case, apart from refuting the very specific hypotheses tested. The overarching question of whether higher-dimensional simplicial structures in neural networks are linked to some form of function, either causal of not, thus remains unanswered. The main reason for this is the problematic vagueness of the hypothesis - it is simply impossible to falsify whether the simplicies serve "some functional role", it can only be verified in the case where one simply happens to test for the correct function. This is a good reason to be critical of such an approach in science.

The final research question was whether the simplices of different sizes form coherent functional groups as well as structural groups, as was also discussed using the notion of natural kinds. The main finding against this grouping was the general lack of within-group homogeneity for all the different functional metrics tested. This was shown by the high within-group variance, and evidenced further by the fact that the analysis of variance in very few of the cases showed that the groupings according to simplex size or completeness were statistically significant. The same problem of choice of function applies here, where there is no guarantee that the "correct" function has been tested for, so this is only an *indication* that these simplices do not form functionally coherent groups. This finding is further supported by the fact that the variance was greatly reduced when the type of source neuron (excitatory or inhibitory) was selected for. This shows that there are other factors than simplex size which greatly affects the functions of the simplices, such that blindly grouping them together purely based on their topology in this manner is not particularly informative of their function. However, as before, the very nature of the question makes it impossible to falsify. In conclusion, the results here obtained are insufficient to give any conclusive answers to either of the research questions considered in this project.

6.2 Limitations

There were a range of other factors limiting the potential scope of this research. We have primarily looked at dyadic relationships within the simplicial structures, in particular causal and statistical relationships between the activity of the source and the sink neurons in the simplices. Yet it is far from certain that the functionality which the simplices facilitate is observable in terms of these metrics. Indeed, much of the current approach to neuroscience is focused around more global properties of bigger networks, which can capture emergent properties of the network dynamics which are not reducible to dynamics at lower scales, such as the dyadic relationships investigated here. We mainly focused on the source and the sink neurons, both because these could be treated as the informational bottlenecks in the simplicial structures, but also because some level of simplification was needed. Yet, it might well be the case that the function we are searching for is more strongly connected with e.g. the dynamics of the mediator-neurons.

One must also not forget that the networks studied are highly idealised, and far from realistic, given that no such structures exist in isolation. It might be that the functionality of the simplicial structures only becomes visible when these structures form part of bigger integrated networks. This is a point commonly raised in the literature regarding the need for new approaches to understanding and analysing the vast amounts of brain data available [54, 59]. The approach here taken was to look at the function in terms of local relations within smaller circuits, but we know that not all properties visible at the global scale are reducible to such low-level properties and interactions [2]. Rather, these are properties which might emerge as a result of the collective interactions of elements in complex systems. Regarding this alternative approach to network neuroscience, Sporns writes:

> In the case of large networks of neurons, powerful manifestations of "emergent phenomena" are global states of brain dynamics in which large populations of neurons engage in coherent and collective behavior. Such dynamics emerge from very large numbers of local interactions that are individually weak and yet collectively powerful enough to create large-scale patterns. [54, p. 92-93]

Thus, the current findings are insufficient to say anything about the significance of simplices for such higher-level emergent properties.

Related to this, one could imagine that the function of these structures only arise in the neuromodulatory environments in which they normally exist, due to some more complex interdependencies. This would be impossible to discover with the current approach. However, in addition to providing a necessary simplification, the idealisation can nevertheless be useful. Even if the true function is different in the full neuromodulatory environment, knowing how the system works in isolation is still valuable. This is similar to how physicists assume frictionless surfaces and no air resistance. The calculations can then give an upper limit for e.g. the movement of an object, to which other factors can be added later. Experimentally, it also provides a way of singling out the effect of different factors. Thus, if we could find ways of more closely recording the structure and activity in such networks *in vivo*, we might be able to get around some of the circularity in the argument outlined in section 2.1.2, where the aim is to discover the dependencies of function on network structure and the neuromodulatory environment.

Next, it is an important limitation that the information in the networks studied remains uninterpreted. When talking of function in biological neural networks, we are primarily interested in higher-order cognitive functions, and the uninterpreted sequences of zeros and ones here investigated are very far from these. Relatedly, we mainly focused on how individual and isolated spikes in the source neurons affected the dynamics of the other neurons in the network. Though, as discussed, no information is encoded in such isolated spikes - it is in the spike series that information exists. Thus, this simplification might have limited the possibilities of detecting the functionalities searched for.

Finally, one must not forget that all the datasets investigated were *simulated* data, so one must be careful with extrapolating the findings from this research to also make claims about biological neural networks. To do so, additional experimental data from real neural networks will be necessary.

6.3 Future Research

We have an ever increasing amount of brain-related "big data" available, though neuroscience still lacks the organising principles and theoretical framework needed to turn this data into new knowledge and understanding. Thus, in addition to new experimental methods and tools for analysing the data, the field is also in dire need of theoretical advances which can help us interpret the data. Neuroscience is by no means unique in this regard, and historically, other sciences such as physics and chemistry have all gone through similar processes where new experimental techniques and theories are developed in tandem until a final agreed upon theoretical framework is reached. Until then, it is unreasonable to judge the methods used in neuroscience by the same standards as one would in for instance physics, but conversely, the models derived from them must be treated with a certain care and reservation. What one must encourage, then, is a mixture of creative and critical voices in the field, where new methods and frameworks are being suggested, tested and scrutinised. The shift in focus from correlation towards causation might hopefully provide one such fruitful change in perspective. It can certainly be argued that causal explanations provide both a stronger link to the underlying causal mechanisms which we wish to discover, as well as having a much clearer interpretation than mere correlations do, so should be more conducive to scientific explanations.

Further, we would like to raise two points of concern which has kept reappearing throughout the work on this project. The first concerns the difficulty of choosing the right level of analysis for features of a complex system, the second regards the pursuitworthiness of the simplicial structures, and other similarly motivated structures.

6.3.1 Levels of Description

The challenge of analysing complex systems, such as the brain, at the right level of description, is one that has been mentioned repeatedly throughout the discussions in this project. This is a problem which appears in all complex systems, and especially in cases such as the brain, where the causal dependencies between the levels likely work in both directions, such that properties emerging at certain higher levels might in turn causally influence the properties at lower levels [9]. Carandini argues that what neuroscience largely lacks is theories

of how the intermediary levels relates to both the underlying biophysics as well as the emergent functionality of behaviour and other cognitive functions. This corresponds well with Marr's computational level, where a complete understanding requires knowledge of how the algorithms at one level relate to the higher computational purpose of the system. What we need are bridge principles which can be used to connect these.

Further, there is a crucial difference between reductionism and constructivism, where it is often tacitly assumed that the former implies the latter, when this is not necessarily the case. This is not unique to neuroscience, but is something we find in complex systems in all fields. Writing about similar challenges in physics and chemistry, Andersen poignantly notes that "[t]he ability to reduce everything to simple fundamental laws does not imply the ability to start from these laws and reconstruct the universe" [2, p. 393]. This concern is generally ignored in approaches such as the current project, where we try to recreate higher-level functionalities from simulations of lower-level dynamics. Regarding this approach in neuroscience, Carandini writes: "In essence, while we have clear examples of success for the reductionist approach (from behavior to computations to circuits), the case still needs to be made for the constructivists' one (from circuits to computations to behavior)" [9, p. 183]. Thus, we have no guarantee that the function searched for will even appear in such simulations, since this could be dependent on factors appearing at higher levels. Indeed, some argue that we cannot even consider the brain in isolation from the body, and that it could well be the case that features such as cognition only appear in the brain as an embodied system [49].

Thus, without knowing what the correct level of analysis for the function of interest is, nor the mapping between the functionalities at the different levels, one is at high risk of not seeing the wood for the trees. If the aim is to test whether certain circuit motifs are related to certain behaviours, it is necessary to study this through the intermediate level of computational purpose. Further, one must be clear on what direction one is working in (reductionism or constructivism), since neither implies the other, and the necessary bridge principles should have been established in the right direction beforehand. This is something which was missing from the current project, where nothing was known about how the functionalities searched for at the level of circuits really connected to any higher-level properties of interest. Thus, even if a positive result had been obtained, we would have been unable to interpret its significance. This is unfortunately not unique to this project, since the lack of known bridge principles is a general problem in the field. Thus, this somewhat speculative approach will likely prevail out of sheer necessity in the near future, so we can only encourage a high degree of scepticism and scrutiny of one's own results.

6.3.2 Pursuitworthiness

The second concern is that of how one decides what features of a system to study and put time and money into researching, in other words, how to determine the *pursuitworthiness* of a feature. The current project is a good illustration of this, where, based mainly on their topological neatness, we decided to search for the function of the strictly defined directed simplices. However, we had little idea about what the correct level of description was for their function, so a negative result does not hinder us from simply rephrasing the hypothesis and go searching further at another level or using a different metric.

In general, we should be very wary of *a priori* picking out certain networks structures of interest and then go looking for their functionality. Such an approach, where one goes hunting for function before determining what function one expects, is usually considered dubious scientific practice. Firstly, it goes against the Popperian falsificationist ideal of science, where scientific hypotheses are required to be falsifiable. Secondly, with such an approach, where numerous different types of functions are iteratively tested for, the likelihood of eventually committing a type I error becomes increasingly large, as one is bound to eventually find something which turns up as statistically significant.

At the same time, as mentioned above, neuroscience is in need of new ideas and perspectives, so one should probably be a bit more lenient with the falsifiability here than one would in other, more established, fields. Indeed, there is a very real tension between stubbornly refusing to let go of the belief in the importance of a certain feature, and therefore relentlessly searching for its function at different levels of description (this would constitute what Lakatos would call a "degenerate research program"), and the problem that it is impossible to know *a priori* what the correct level of description for the function of this feature might be. Thus, it would be preferable if any further search for the function of directed simplices in biological networks were supported by additional experimental or theoretical findings.

References

[1] U. Alon. Network motifs: theory and experimental approaches. *Nature Reviews Genetics*, 8(6):450–461, June 2007.

[2] P. W. Anderson. More is different. *Science*, 177, 1972.

[3] C. I. Bargmann. Beyond the connectome: How neuromodulators shape neural circuits. *BioEssays*, 34(6):458–465, June 2012.

[4] C. I. Bargmann and E. Marder. From the connectome to brain function. *Nature Methods*, 10(6):483–490, June 2013.

[5] C. M. Bennett, M. B. Miller, and G. L. Wolford. Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: an argument for multiple comparisons correction. *NeuroImage*, 47:S125, July 2009.

[6] A. Bird and E. Tobin. Natural Kinds. In Edward N. Zalta and Uri Nodelman, editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Spring 2023 edition, 2023.

[7] F. J. Boge. Two Dimensions of Opacity and the Deep Learning Predicament. *Minds and Machines*, 32(1):43–75, March 2022.

[8] H. Brunborg and J. Sønstebø. Neuro ml summer project, 2022. URL: `https://gitlab.com/hermabr/neuro_ml`.

[9] M. Carandini. From circuits to behavior: A bridge too far? In Gary Marcus and Jeremy Freeman, editors, *The Future of the brain : essays by the world's leading neuroscientists*, pages 177–185. Princeton University Press, Princeton, N.J, 2015.

[10] N. Cartwright. *How the laws of physics lie*. Clarendon Press, Oxford, 1983.

[11] M. Chalfie, J. E. Sulston, J. G. White, E. Southgate, J. N. Thomson, and S. Brenner. The neural circuit for touch sensitivity in caenorhabditis elegans. *Journal of Neuroscience*, 5, 1985.

[12] M. Cobb. *The idea of the brain : a history*. Profile Books, London, England, 2021.

[13] C. F. Craver. The explanatory power of network models. *Philosophy of Science*, 83(5):698–709, 2016.

[14] C. Curto and V. Itskov. Cell groups reveal structure of stimulus space. *PLoS Computational Biology*, 4, 2008.

[15] C. P. Doncaster and A. J.H. Davey. *Analysis of variance and covariance: How to choose and construct models for the life sciences*. Cambridge University Press, 2007.

[16] M. Eisenstein. Artificial intelligence powers protein-folding predictions. *Nature (London)*, 599(7886):706–708, 2021.

[17] J. S. Elam, M. F. Glasser, M. P. Harms, S. N. Sotiropoulos, J. L.R. Andersson, G. C. Burgess, S. W. Curtiss, R. Oostenveld, L. J. Larson-Prior, J. Schoffelen, M. R. Hodge, E. A. Cler, D. M. Marcus, D. M. Barch, E. Yacoub, S. M. Smith, K. Ugurbil, and D. C. Van Essen. The Human Connectome Project: A retrospective. *NeuroImage*, 244:118543, December 2021.

[18] C. Z Elgin. *True enough*. MIT Press, Cambridge, MA, 2017.

[19] A. Fornito, A. Zalesky, and E. T. Bullmore. *Fundamentals of Brain Network Analysis*. Academic Press, 2016.

[20] W. Gerstner, W. M. Kistler, R. Naud, and L. Paninski. *Neuronal Dynamics: From Single Neurons to Networks and Models of Cognition*. Cambridge University Press, USA, 2014.

[21] C. Giusti, R. Ghrist, and D. S. Bassett. Two's company, three (or more) is a simplex: Algebraic-topological tools for understanding higher-order structure in neural data. *Journal of Computational Neuroscience*, 41(1):1–14, August 2016.

[22] S. J. Gould and R. C. Lewontin. The spandrels of san marco and the panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society of London - Biological Sciences*, 205, 1979.

[23] D. R. Whitney H. B. Mann. On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 18, 1947.

[24] K. He, X. Zhang, S. Ren, and J. Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015.

[25] C. C. Hilgetag and M. Kaiser. Clustered organization of cortical connectivity. *Neuroinformatics*, 2, 2004.

[26] H. K. Inagaki, S. Ben-Tabou De-Leon, A. M. Wong, S. Jagadish, H. Ishimoto, G. Barnea, T. Kitamoto, R. Axel, and D. J. Anderson. Visualizing neuromodulation in vivo: Tango-mapping of dopamine signaling reveals appetite control of sugar sensing. *Cell*, 148, 2012.

[27] S. P. Jensen. Edge regressor for simplicial complexes, 2023. URL: `https://github.com/SaraPJensen/Git_Master/tree/main/simplex_complete`.

[28] S. P. Jensen. Neural activity simulator, 2023. URL: `https://github.com/SaraPJensen/snn-glm-simulator/tree/sara_simplex`.

[29] E. Kim, H. Kang, H. Lee, H. J. Lee, M. W. Suh, J. J. Song, S. H. Oh, and D. S. Lee. Morphological brain network assessed using graph theory and network filtration in deaf adults. *Hearing Research*, 315, 2014.

[30] W. H. Kruskal and W. A. Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47, 1952.

[31] H. Lee, M. K. Chung, H. Kang, B. N. Kim, and D. S. Lee. Discriminative persistent homology of brain networks. *Proceedings - International Symposium on Biomedical Imaging*, 2011.

[32] M. E. Lepperød, T. Stöber, T. Hafting, M. Fyhn, and K. P. Kording. Inferring causal connectivity from pairwise recordings and optogenetics. *bioRxiv*, 2022.

[33] H. Levene. Robust tests for equality of variances. *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, 69, 1960.

[34] N. K. Logothetis. What we can do and what we cannot do with fMRI. *Nature*, 453(7197):869–878, June 2008.

[35] E. Marder. Neuromodulation of neuronal circuits: Back to the future. *Neuron*, 76, 2012.

[36] D. Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information.* Henry Holt and Co., Inc., New York, NY, USA, 1982.

[37] J. D. Medaglia, M. E. Lynall, and D. S. Bassett. Cognitive network neuroscience. *Journal of Cognitive Neuroscience*, 27:1471–1491, 8 2015.

[38] J. L. Morgan and J. W Lichtman. Why not connectomics? *Nature Methods*, 10(6):494–500, June 2013.

[39] K. Morrison and C. Curto. Predicting neural network dynamics via graphical analysis. 2018.

[40] A. Motta, M. Berning, K. M. Boergens, B. Staffler, M. Beining, S. Loomba, P. Hennig, H. Wissler, and M. Helmstaedter. Dense connectomic reconstruction in layer 4 of the somatosensory cortex. *Science*, 366, 2019.

[41] V. N. Murthy. Synaptic plasticity: Step-wise strengthening. *Current Biology*, 8, 1998.

[42] I. Neutelings. Neural networks. URL: `https://tikz.net/neural_networks/`.

[43] M. A. Nielsen. Neural networks and deep learning, 2018. URL: `http://neuralnetworksanddeeplearning.com/`.

[44] L. Paninski, J. W. Pillow, and E. P. Simoncelli. Maximum likelihood estimation of a stochastic integrate-and-fire neural encoding model. *Neural Computation*, 16, 2004.

[45] J. Pearl. *Causality: Models, reasoning, and inference, second edition.* Cambridge University Press, 2011.

[46] J. W. Pillow, J. Shlens, L. Paninski, A. Sher, A. M. Litke, E. J. Chichilnisky, and E. P. Simoncelli. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207):995–999, 2008.

[47] M. W. Reimann, M. Nolte, M. Scolamiero, K. Turner, R. Perin, G. Chindemi, P. Dłotko, R. Levi, K. Hess, and H. Markram. Cliques of Neurons Bound into Cavities Provide a Missing Link between Structure and Function. *Frontiers in Computational Neuroscience*, 11:48, June 2017.

[48] T. Schreiber. Measuring information transfer. *Physical Review Letters*, 85, 2000.

[49] L. Shapiro and S. Spaulding. Embodied Cognition. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2021 edition, 2021. URL: `https://plato.stanford.edu/archives/win2021/entries/embodied-cognition/`.

[50] S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52, 1965.

[51] O. Sporns. *Networks of the brain*. MIT Press, Cambridge, Mass, 2011.

[52] O. Sporns. Making sense of brain network data. *Nature Methods*, 10(6):491–493, June 2013.

[53] O. Sporns. Contributions and challenges for network models in cognitive neuroscience. *Nature Neuroscience*, 17, 2014.

[54] O. Sporns. Network neuroscience. In Gary Marcus and Jeremy Freeman, editors, *The Future of the brain : essays by the world's leading neuroscientists*, pages 90–100. Princeton University Press, Princeton, N.J, 2015.

[55] C. J. Stam. Modern network science of neurological disorders. *Nature Reviews Neuroscience*, 15, 2014.

[56] A. O.W. Stretton, R. E. Davis, J. D. Angstadt, J. E. Donmoyer, and C. D. Johnson. Neural control of behaviour in ascaris. *Trends in Neurosciences*, 8(C):294–300, 1985.

[57] M. F. Triola. *Elementary statistics technology update*. Pearson Education, Boston, 11th ed edition, 2012.

[58] J. R. Turner. *Introduction to analysis of variance : design, analysis, & interpretation*. SAGE, Thousand Oaks, Calif. ;, 2001.

[59] K. v. Shenoy. Recording from many neurons simultaneously - from measurement to meaning. In Gary Marcus and Jeremy Freeman, editors, *The Future of the brain : essays by the world's leading neuroscientists*, pages 78–89. Princeton University Press, Princeton, N.J, 2015.

[60] J. Ward. *The Student's Guide to Cognitive Neuroscience*. Routledge, Taylor & Francis Group, 2019.

[61] B. L. Welch. On the comparison of several mean values: An alternative approach. *Biometrika*, 38, 1951.

[62] A. Wnuk, A. Davis, C. Parks, D. Halber, D. Kelly, G. Zyla, K. Hopkin, K. Weintraub, J. M. Beverly, K. S. Sheikh, L. Wessel, L. Chiu, M. Fessenden, M. Galinato, M. Richardson, S. Blumenrath, and S. Rojahn. *Brain Facts: A Primer on the Brain and Nervous System*. Society for Neuroscience, 2018.