

Stimulating artificial neurons with simplex and torus networks: an information-theoretic study.

Jacob Bamberger

Supervised by Professor Kathryn Hess Bellwald

February 13, 2021

Abstract

This exploratory project aims at studying neural networks using tools from information theory. Motivated by the increasing evidence of a link between structure and function of neural networks (ex [19]), we study the relationship between the graph structure and the dynamics, of certain small neural networks, which were selected for their topological and structural properties. The underlying dream that guided our thought process and experimentation was to be able to use spiking data from many simulations of a fixed network to reconstruct the graph corresponding to the network. We therefore first construct a probabilistic model from spiking data of many simulations of some neural network. We then analyze and interpret information measures such as the entropy of a neuron, and mutual information between neurons as well as the impact of conditioning on these measures. We find that through the lens of information theory, the probabilistic model is very relevant as it pertains to the structure of networks. One type of network that stands out in this respect is the high dimensional directed simplex of neuron, which allows for a wide range of neuronal behaviour, while being a reliable network to communicate information between neurons.

Contents

1	Introduction	2
2	Background material	3
2.1	The Simple Model of choice	3
2.2	Information Theory	4
2.2.1	Entropy	5
2.2.2	Mutual Information	6
2.2.3	Channel Capacity	7
2.2.4	Multivariate Measures	8

2.2.5	The Information Simplex	10
2.3	Topology and Structure of Networks	11
2.3.1	Simplices and simplicial complex	12
2.3.2	Filtered simplex	12
3	The Set Up	13
3.1	The Simple Framework of Choice	13
3.2	Generating and Measuring Observations	14
3.2.1	From topology to network	14
3.2.2	From network to spikes	14
3.2.3	From spikes to distributions	14
4	Experiments and Results	16
4.1	Entropy and number of parents	16
4.2	Mutual information and number of common parents	17
4.2.1	Mutual information of two neurons	18
4.2.2	Mutual information and common parent network	19
4.2.3	Adding connections between children	20
4.3	Conditional entropy, conditional mutual information, and predictability of neurons.	22
4.3.1	Conditional entropy and neuron predictability	22
4.3.2	Bell shaped conditional mutual information	22
4.3.3	Towards an information gradient	24
4.3.4	Disconnected neurons are not independent	26
4.3.5	Integrating the bell shaped curve.	27
4.4	Three Neuron Motifs	28
5	Further Directions	29
5.1	More diverse Population of Neurons and Synapses	29
5.2	Other Measurements	29
5.3	Other networks	29
5.4	Capacity and Structure of Channel	31
5.5	Applications to network classification	31
5.6	Information gradient and the high dimensional simplex	32
6	Conclusion	32
7	Appendix	36

1 Introduction

This study is inspired, and can in some sense is an extension of Skander Moalla's bachelor thesis [17]. The methodology is shared, but we focus on different phenomena, so this report assumes no prior knowledge of the other, although we refer to it for more details on certain aspects. These projects are inspired by [19], in which the authors combine topological and statistical ideas to study the

relationship between function and structure of neural networks, [18]. In recent years, information theory has found many applications in neuroscience, we thus develop further insight on the relationship between structure and function of neural networks by using tools from information theory.

We first introduce several-information theoretic concepts, and a process to build a probabilistic model of spiking neuron data for given networks. We apply the information theoretic tools to the probabilistic model, shedding light on the relation between structure and function of neural networks. The networks considered are organized as a group of parent neurons and a group of child neurons, where each parent has a connection to each of the children. Further structure is studied among the parent neurons, as well as among the children neurons.

In Section 2, we introduce the theoretical background we use in the study. We first introduce the neuroscientific models used to generate the data. We then introduce several information-theoretic tools. Finally we explain what we mean by structure of networks.

In Section 3, we outline how the data are generated using public software, available online and easy to use. We then outline how the data is generated and how the probabilistic models are constructed.

In Section 4, we run and analyze experiments. First we examine how the randomness of a neuron relates to the network it lives in. Then we measure the influence the ambient network has on how similarly two neurons can behave. We then study how the behaviour of group of neurons can inform us on a specific neuron. Finally, we use the information simplex to do an exhaustive study to describe the information properties of all networks on three neurons without self loops.

In Section 5, we outline directions we would have pursued if the duration of a semester were unlimited...

2 Background material

2.1 The Simple Model of choice

In this section we describe the neuroscientific model used for our study. The most important objects of interest when studying the brain are neurons and synapses. Our model will therefore consider only these two objects. There are several ways to describe neurons as a dynamical system. Several models have been defined and studied from both a mathematical and a biological perspective. Some well known models, such as the "integrate and fire" model [13], have been criticized for being overly simple and not biologically accurate. Other options such as the Hodgkin-Huxley-type models [13] are neuro-computationally interesting but computationally too expensive to run on a personal computer. We therefore seek a middle ground between computational efficiency and biological accuracy.

This should be provided by the "simple model of choice". We refer the reader to [12] for a biologically inclined study and to [13] for a mathematical analysis.

The model can be described as a two dimensional system of ordinary differential equations:

$$\begin{aligned} v' &= 0.04v^2 + 5v + 140 - u + I \\ u' &= a(bv - u) \end{aligned}$$

together with an auxiliary after-spike resetting: if $v \geq 30$ mV, then

$$\begin{aligned} v &\leftarrow c \\ u &\leftarrow u + d, \end{aligned}$$

where a, b, c, d are constant parameters, and u, v, I are functions of time. The variable v represents the membrane potential of the neuron. The variable I represents the incoming current. The variable u represents membrane recovery. As one expects, a positive stimulus I will produce a positive change in the membrane potential v , and in contrast a larger value of the membrane recovery u will decrease the excitability. The auxiliary after-spike resetting equations gives the threshold (30mV) for neuron spiking and gives new initial conditions for v and u .

Another representation of this dynamical system is the following system of differential equations:

$$\begin{aligned} Cv' &= k(v - v_r)(v - v_t) + u - I \\ u' &= a(b(v - v_r) - u) \\ \text{if } v &\geq v_{peak} \text{ then} \\ v &\leftarrow c, u \leftarrow u + d. \end{aligned}$$

Here v is the membrane potential, u is the recovery current, C is the membrane capacitance, v_r is the resting membrane potential, v_t is the instantaneous threshold potential, and v_{peak} is the spiking threshold. For more details we refer the reader to [13].

The synapses are modeled with an instantaneous change in the membrane potential of the post-synaptic neuron. This means that when the pre-synaptic neuron fires, the synapse outputs a constant charge to the post-synaptic neuron. We refer to the value of the charge as the "synapse weight".

2.2 Information Theory

In this section we give an overview of the information theoretic preliminaries for this study, with a view towards experimental sciences and particularly neuroscience. Modern information theory is often attributed to Shannon for its first formalization in [20]. The seminal work was born out of a need for a general theory of communication.

It is natural to apply these tools to neuroscience and particularly to neural networks as one of their main function is to store and to process information. One may hope that using this language can enable us to formulate questions and perhaps even provide answers about the link between structure and function of biological neural networks.

For more detailed introductory notes and explanations of classical information theory, namely ones presented in Sections 2.2.1, 2.2.2 & 2.2.3 we refer the reader to [6]. To compute distributions and information theoretical measures we have relied on the "discrete information information" software [14], the documentation of the software includes a concise introduction to the measures presented in this section, as well as examples. The figures in Section 2.2.4 come from there.

2.2.1 Entropy

One fundamental concept in information theory is that of entropy.

When studying a system, one may make experiments and get observations, shedding light on how the system behaves. Each observation will provide "information" about the state of the system. The concept of entropy can be thought of as a measure of the number of observations one has to make to know the systems behaviour. We call the "entropy" of a system the amount of randomness such a system contains. Just like a simple yes/no observation represents one bit of information, this quantity is measured in bits.

Our system will be formalized as a discrete random variable X that takes values in a finite set $\Omega = \{x_1, x_2, \dots, x_n\}$. The *probability mass function* of X is the function $p(X = x)$ or $p_X(x)$ which should be read "*the probability that X takes value x is ...*". The entropy of X is then

$$H(X) = - \sum_{x \in \Omega} p_X(x) \log_2(p_X(x)).$$

Notice that given two random variables X and Y taking values in Ω_X and Ω_Y one may combine them to get a new 'vector' random variable $Z = (X, Y)$ taking values in $\Omega_X \times \Omega_Y$ with probability mass function p_Z equal to the joint distribution of X and Y , $p(x, y)$ enabling us to define the *joint entropy* of two random variables as

$$\begin{aligned} H(X, Y) &= - \sum_{(x,y) \in \Omega_X \times \Omega_Y} p(X = x, Y = y) \log_2(p(x, y)) \\ &= - \sum_{z \in \Omega_Z} p_Z(z) \log_2(p_Z(z)) = H(Z). \end{aligned}$$

As we are interested in analyzing the interaction of several neurons to investigate network connections, we will also need a way to look at the impact of one variable on another. How random is a random variable if you already know the value of another random variable? A measurement can be seen as a random variable if the same measurement is performed on many different samples, since the value of the measurement will vary depending on the state of the system.

Conversely, the measurement will provide information on the state of the system. The post-measurement randomness is exactly what conditional entropy captures.

For two random variables X and M we define the *entropy of X conditioned on M* as follows:

$$\begin{aligned} H(X|M) &= - \sum_{m \in \Omega_M} p_M(m) H(X|M = m) \\ &= - \sum_{m \in \Omega_M} p_M(m) \sum_{x \in \Omega_X} p(X = x|M = m) \log_2(p(X = x|M = m)). \end{aligned}$$

In the first line, $H(X|M = m)$ is the entropy of X knowing that $M = m$ and can be interpreted as the randomness of X knowing that the measurement of M is m . Therefore, conditional entropy can be understood as the average randomness over all measurements. Most information theoretical measures have a conditional version, which is very interesting in terms of interpretability, and of the relation between random variables.

2.2.2 Mutual Information

Building on the notion of conditional entropy, which is a measurement of how random a system is knowing the value of some observation concerning this system, we now formalize the idea of how much information a variable contains about another. As hinted previously, an observation about the system can only decrease the randomness of that system, therefore one candidate for how much one random variable informs us about another is the decrease in randomness from before the measurement to after the measurement.

Indeed the *mutual information between two random variables X and Y* is defined as follows:

$$\begin{aligned} I(X; Y) &= H(X) - H(X|Y) \\ &= H(X) + H(Y) - H(X, Y) \end{aligned}$$

We first point out the symmetry of the second line which is important in quantifying shared information. One way to visualize this quantity is in terms of Venn diagrams, as in Figure 1. The union of the two sets is the joint entropy, the red and blue areas are the entropies of each individual random variables, the red area without the intersection is the conditional entropy and the intersection is the mutual information. In view of this, one can think of the mutual information as the randomness shared by both variables.

As for the entropy, we may replace a random variable by a vector of a random variables (which is again a random variable!). This allows us to compute the quantity $I((X, Y); W)$ which is the mutual information between the vector

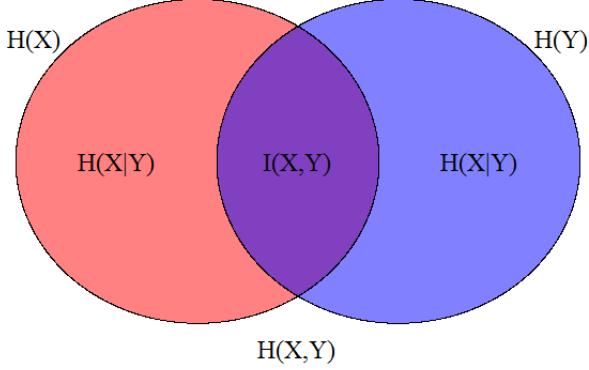


Figure 1: Visualization of the joint entropy, entropies, conditional entropies and mutual information of two random variables. Obtained from [23].

(X, Y) and W , enabling us to analyze how two random variables work together to constrain a third random variable.

Every information measure has its conditional counter-part, and mutual information is no exception. Indeed one may ask how much information variable X gives about variable Y if we already know the value of variable M . For two different values of M , there may be more or less mutual information between X and Y . One way of viewing the conditional mutual information is as an average over all values of M . Another way to view it is as the difference between the randomness of X knowing M and the randomness of X knowing both M and Y ; i.e. how much uncertainty is removed by knowing both M and Y versus knowing only M . These two perspectives are summarized in the following equation.

$$\begin{aligned}
 I(X; Y|M) &= - \sum_{m \in \Omega_M} p_M(m) I(X; Y|M = m) \\
 &= - \sum_{m \in \Omega_M} p_M(m) (H(X|M = m) + H(Y|M = m) - H(X, Y|M = m)) \\
 &= H(X|M) - H(X|(M, Y))
 \end{aligned}$$

2.2.3 Channel Capacity

The formalization of the notions of communication channel, and channel capacity, and the practical interpretation thereof is the central success of information theory.

A communication channel is the means by which a sender, A, can send a message to a receiver, B, so that message received by B is the same as the message that was sent by A. A given channel can allow for a certain number of distinguishable messages, expressed in terms of a logarithm: the number of bits the channel can send viably. We formalize the latter below as the channel capacity, indeed the celebrated *channel coding theorem* states that the formalization below is a precise measurement of the efficiency of a given channel: it is the maximum number of bits per channel use one can send viably.

Since a communication channel is in general not deterministic, probabilistic language is therefore appropriate. A message can be thought of as a random variable X , which takes value in a set of messages Ω_X . Sending this message through a fixed communication channel will produce a new message Y taking values in Ω_Y (for simplicity of the exposition we will assume that $\Omega_X = \Omega_Y$ and that the encoding and decoding functions are the identity). It is clear that the output of the communication channel will depend on the value of the initial message X . It is therefore natural to think of the communication channel as a conditional probability density function $p(Y = y|X = x)$. Notice that in this context it is necessary and sufficient to define the probability density function p_X in order to get a probability density function $p_Y(y) := \sum_{x \in \Omega_X} p_X(x)p(Y = y|X = x)$. One may thus see the mutual information between X and Y , $I(X; Y)$ as a certain similarity measure between the sent message and the received message. Indeed the channel capacity C of a (discrete, memoryless) channel is defined as:

$$C = \max_{p_X} I(X; Y)$$

where the maximum is taken over all distributions p_X on Ω_X . Note that C is indeed only a function of $p(Y|X)$.

2.2.4 Multivariate Measures

Information theory, particularly the mutual information measure, has had great success in formalizing the similarity and interaction between two random variables. It is therefore natural to ask if a generalization of this measure will be as successful in formalizing the interactions among three or more variables. Several measures have been introduced, each claiming to measure a version of multi-information. In this section we list and define several of these measures. A more thorough study of these can be found in [].

Recall that the total randomness for a set of variable $\{X_1, \dots, X_n\}$ is simply the joint entropy $H(X_1, \dots, X_n)$ of all the random variables.

The following measures all generalize mutual information, as they are equal to it when the number of random variable is equal to two.

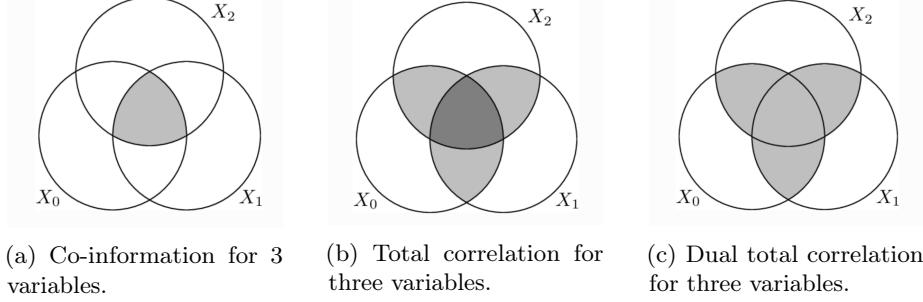


Figure 2: Multivariate mutual information measures. Obtained from [14]

The *co-information* of a set of random variables was defined in [4] as

$$I_{co}(X_1, X_2, \dots, X_n) := \sum_{S \subset \mathcal{P}} (-1)^{|S|} H(X_S), \text{ where } X_S = \{X_i : i \in S\},$$

where \mathcal{P} is the power set of $\{1, 2, \dots, n\}$.

Just like mutual information, this has a nice interpretation in terms of Venn diagrams, as seen in Figure 2a, where for each intersection we get a type of mutual information. This co-information represents the quantity at the center of the diagram. It is suggested that a non-zero value corresponds to a n-th order dependency. These values can however be negative and there is no straightforward interpretation of what such a phenomenon means.

The second multivariate information measure we consider is the *total correlation* defined in [22] as

The interpretation in terms of Venn diagrams for $n = 3$ can bee seen in Figure 2b. It corresponds to the union of all possible intersections between 2 sets plus the information of the intersection of the 3 sets. The natural way to look at this is the following: it is the difference between the entropy of all individual random variables, and the entropy of the whole. It is zero if every random variable were independent. Indeed this quantity is equal to the Kullback-Leibler divergence between the joint distribution and the distribution where all random variables are independent. This is a non-negative measure, and is equal to zero if and only if the variables are independant.

The third multivariate measure we consider is the *dual total correlation* [9], defined as:

$$B(X_{1\dots n}) = H(X_{1\dots n}) - \sum_{i=1}^n H(X_i | X_{\{1\dots n\} - \{i\}}).$$

Its interpretation in terms of Venn diagrams can be seen in Figure 2c.

2.2.5 The Information Simplex

In this section we describe an information theoretic summary of interactions among several random variables. These are observations we have made that we apply in Section 4.4 to construct the information 3-simplex of several 3 neuron networks. We would be very interested in doing a longer study of these techniques, and potentially to relate them to the fields of information geometry [1] and information topology [3].

We first refer the reader to Sections 2.3.1 & 2.3.2 for definitions of the n -simplex and filtration thereof. From a set of random variables $\{X_0, X_1, \dots, X_n\}$ with joint probability distribution $p_{\{0,1,\dots,n\}}$, we construct a filtration of the n -simplex as follows. Define the function H on faces of the n -simplex to be the joint entropy of the random variables indexed by the vertices of the given face, i.e. $H([a_0, a_1, \dots, a_i]) := H(X_{a_0}, X_{a_1}, \dots, X_{a_i})$. Since the value of H is larger on a face than on any of its sub-faces, we may consider the associated filtration of the simplex. We now show how this simplex can be used to obtain several of the previously defined information measures.

A theoretical insight is that the co-information measure of a set of random variables is equal to the integral of the function H against the Euler characteristic over the face corresponding to the set of random variables.

$$I_{co}(X_{a_0}; X_{a_1}; \dots; X_{a_i}) = - \int_{[a_0, a_1, \dots, a_i]} H d\chi.$$

Furthermore, conditioning on a set of random variables yields a further interpretation. We state the result in a special case of four random variables, as the general formula has yet to be proved:

$$I_{co}(X_0; X_1; X_2 | X_3) = \int_{[0,1,2]} H d\chi - \int_{[0,1,2,3]} H d\chi.$$

Proof.

$$\begin{aligned}
I_{co}(X_0; X_1; X_2|X_3) &= H(X_0; X_1; X_2|X_3) \\
&\quad - H(X_0; X_1|X_3) - H(X_0; X_2|X_3) - H(X_1; X_2|X_3) \\
&\quad + H(X_0|X_3) + H(X_1|X_3) + H(X_2|X_3) \\
&= H(X_0; X_1; X_2; X_3) - H(X_3) \\
&\quad - H(X_0; X_1; X_3) + H(X_3) - H(X_0; X_2; X_3) + H(X_3) - H(X_1; X_2; X_3) + H(X_3) \\
&\quad + H(X_0; X_3) - H(X_3) + H(X_1; X_3) - H(X_3) + H(X_2; X_3) - H(X_3) \\
&= H(X_0; X_1; X_2; X_3) - H(X_0; X_1; X_3) - H(X_0; X_2; X_3) - H(X_1; X_2; X_3) \\
&\quad + H(X_0; X_3) + H(X_1; X_3) + H(X_2; X_3) - H(X_3) \\
&= H(X_0; X_1; X_2; X_3) \\
&\quad - H(X_0; X_1; X_3) - H(X_0; X_2; X_3) - H(X_1; X_2; X_3) - H(X_0; X_1; X_2) \\
&\quad + H(X_0; X_3) + H(X_1; X_3) + H(X_2; X_3) + H(X_1; X_2) + H(X_0; X_1) + H(X_0; X_2) \\
&\quad - H(X_3) - H(X_0) - H(X_1) - H(X_2) \\
&\quad + H(X_0; X_1; X_2) - H(X_1; X_2) - H(X_0; X_1) - H(X_0; X_2) + H(X_0) + H(X_1) + H(X_2) \\
&= - \int_{[0,1,2,3]} H d\chi + \int_{[0,1,2]} H d\chi
\end{aligned}$$

□

The previous equation can be interpreted as integrating over all faces of the simplex $[0, 1, 2, 3]$ that are incident to [3].

Our guess is that in general:

$$I_{co}(S_0; S_2; \dots; S_{i-1}|S_i) = \int_{[\bigcup_{j=0}^{j=i-1} S_j]} H d\chi - \int_{[\bigcup_{j=0}^{j=i} S_j]} H d\chi,$$

where $S_i \subset \{0, 1, \dots, n\}$.

Lastly, we state the **chain rule for co-information**:

$$I_{co}(X_0, X_3; X_1; X_2) = I_{co}(X_0; X_1; X_2) + I_{co}(X_3; X_1; X_2|X_0)$$

2.3 Topology and Structure of Networks

We formalize the notion of structure in neural networks using graph theory and formulate structural complexity using topological and graph-theoretic concepts. We start by abstracting the neural network as a directed graph, where nodes are neurons and directed edges are synapses directed from their pre-synaptic neuron to post-synaptic neuron. Of course, all information is lost in such an abstraction, all neurons are considered equivalent and all synapses as well. One could label the neurons with model parameters to have a finer abstraction, but it is unclear to us how to formulate complexity taking into account these parameters, so this direction has not been pursued.

There are several ways to define structural complexity of a graph, one can look at the total number of edges, the density of the edges or the average in/out degree for each node. These notions are however overly simplistic, in the sense that many different types of graph can have similar complexity measures. Luckily, more robust tools have been developed to study (directed) graphs and their complexity, which we believe have tremendous potential in studying complex systems and particularly neural networks. In this text we focus on tools from algebraic topology, but we believe there are several points of view to explore, namely ones from spectral graph theory, algebraic graph theory, and expander graphs. (see for example [5, 7, 11]).

2.3.1 Simplices and simplicial complex

In order to be able to define topological complexity of a directed graph, we require tools from algebraic topology. Algebraic topology is a broad field of mathematics that concerns the relationship between spaces and algebraic structures. One success of algebraic topology is the classification of spaces using algebraic structures. For a thorough introduction to algebraic topology we refer the reader to [10].

In this study we will restrict ourselves to the study of certain types of spaces, called simplicial complexes, which are higher dimensional analogs of a graph, which are themselves built from even simpler spaces, called simplices.

The building blocks of graphs are the 0-simplex (nodes) and the 1-simplex (vertices). To these building blocks, we can add the 2-simplex, the 3-simplex and even the n -simplex for any natural number n , which will be the dimension of the simplex. The n -simplex exists as the convex hull of n affinely independent points in $(n+1)$ -dimensional Euclidean space. For example, the 2-simplex is a filled triangle, a 3-simplex is a filled tetrahedron. An n -simplex formed by the vertices (or 0-simplices) v_0, \dots, v_n can be denoted by $[v_0, \dots, v_n]$. For this to fit into the directed graph framework, it is convenient to set an ordering of the vertices of such n -simplex, which will thus be denoted by the ordered set $[v_0, \dots, v_n]$. Simplices of dimension up to 3 can be seen in Figure 3.

A simplicial complex is obtained by gluing several simplices along faces. A face of a simplex $[v_0, \dots, v_n]$ is a simplex $[v_{n_0}, \dots, v_{n_i}]$ whose vertices v_{n_j} are also vertices of the original simplex. An example of a simplicial complex that is not a simplex can be seen in Figure 5.

Directed graphs therefore give rise to simplicial complexes naturally, where the n -simplices are in one-to-one correspondence to the list of vertices v_0, \dots, v_n such that (v_i, v_j) is a directed edge from v_i to v_j if and only if $i < j$. The simplicial complex obtained is called the *directed flag complex* associated to the original graph.

2.3.2 Filtered simplex

A *filtered simplex* is a simplex $\sigma := [v_0 \dots v_n]$ with the additional structure, namely a f that associates to each face of σ a positive real number such that for

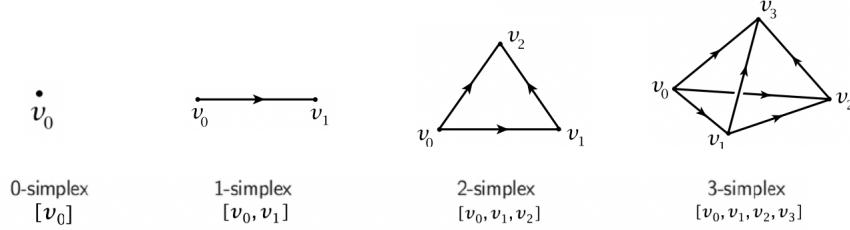


Figure 3: n -simplices with given orderings. Obtained from [10].

two faces ρ and γ of σ , if ρ is also a face of γ then $f(\rho) \leq f(\gamma)$.

3 The Set Up

3.1 The Simple Framework of Choice

We introduce Skander Moalla's [17] implementation of the "simple framework of choice" using the Brian Simulator [8]. The Brian Simulator is a computationally efficient python package that allows for simulating user-defined models of networks of neurons, and simultaneously recording individual neuronal and synapse behaviour.

As the name suggests, the "simple framework of choice" is based on the simple model of choice introduced in Section 2.1. The user can easily define a neural network by inputting a number of neurons, and two lists of equal length where the i -th entry of the first and second lists determine respectively the pre-synaptic and post-synaptic neurons of the i -th synapse. This will generate and simulate a neural network with input connectivity structure and with predefined neuronal and synaptic model. The neurons are modelled using the simple model of choice, and synapses are modeled with an instantaneous change in the membrane potential of the post-synaptic neuron, as described in Section 2.1. Furthermore, each neuron receives a noisy input following a normal distribution.

This allows for great flexibility in the size and structure of the network. The model for neurons and synapses is however fixed. Jupyter notebooks and python code can also be found in the appendix to generate the experiments of this report.

For more details and a more extensive study of the model above we refer the reader to Skander Moalla's Bachelor thesis [17], which can be obtained by contacting us.

3.2 Generating and Measuring Observations

3.2.1 From topology to network

In this section we describe how the specific networks used in the experiment were chosen. Since a choice had to be made of what networks to consider, we let a topological instinct guide us. We were led to several networks that we considered particularly interesting from a structural perspective.

As we have seen in Section 2.3.1, from a directed graph we may construct the directed flag complex, from which we get a topological space. Inversely we may start with a (sufficiently nice) topological space, decompose it into the directed flag complex of some given directed graph (by triangulating the space) which will be the graph we decide to study.

We picked four spaces of particular interest, which we could scale with the number of nodes by keeping their complexity characteristic: the completely disconnected graph, the fully connected graph, the simplex and the torus. These can be seen in Figure 4. Out of these figures, only the torus' structure is hard to interpret, so we add a picture of an undirected triangulated torus in Figure 5.

3.2.2 From network to spikes

Starting from a graph $G = (V, E)$, we describe how the experiments were run. Each node represents a neuron (with a particular set of hyper-parameters), and each directed edge represents a synapse (with a particular set of hyper-parameters) from the parent neuron to the child neuron (from the pre-synaptic to post-synaptic, in usual neuroscientific language).

Once the network is constructed, we simulate the activity in the network and record spontaneous activity, i.e. without any external stimulus, for a total of 1000ms. We divide the total time into time bins of a fixed size, and for each time bin in each simulation we count the total number of spikes for each individual neuron. If, for example, the time bin size is 50ms then we will get 20 different vectors of size equal to the total number of neurons, where the entry at a given row is the number of times the neuron corresponding to that row fired within that time bin. In Figure 6, neuron 0 spikes once in time bin 3, whereas neuron 14 spikes four times in time bin 3.

3.2.3 From spikes to distributions

In order to carry out a statistical or information theoretic study, we need to generate distributions. This means we must collect a large enough amount of data on a system to study. To do so, we fix a network and repeat the previous step many times; for each simulation we get a vector per time bin, these can be computed from raster plots as in Figure 6. We get one distribution per time bin by setting the probability of a vector to appear as the number of times this vector appeared in a simulation divided by the total number of simulations. In Table 3 in the Appendix, we summarize all hyper parameters of our model.

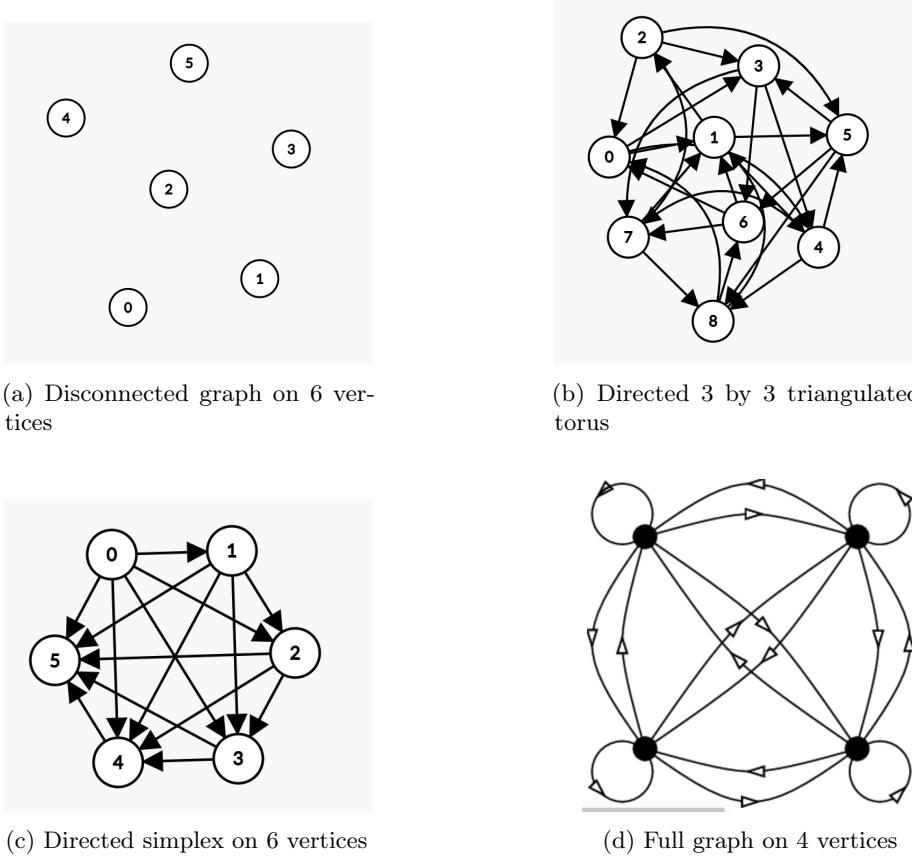


Figure 4: Different networks considered. generated with [?]

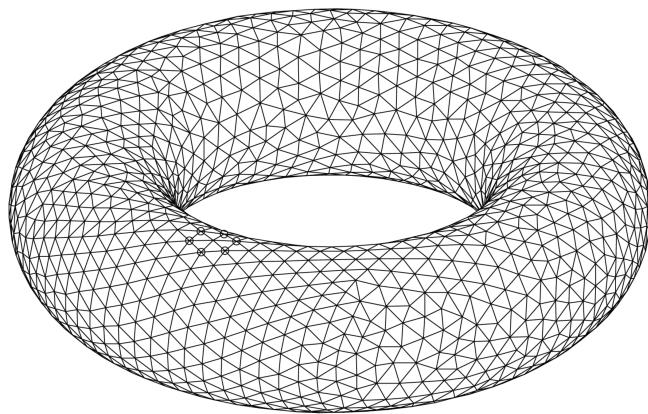


Figure 5: Undirected triangulate torus on many vertices. Obtained from [24]

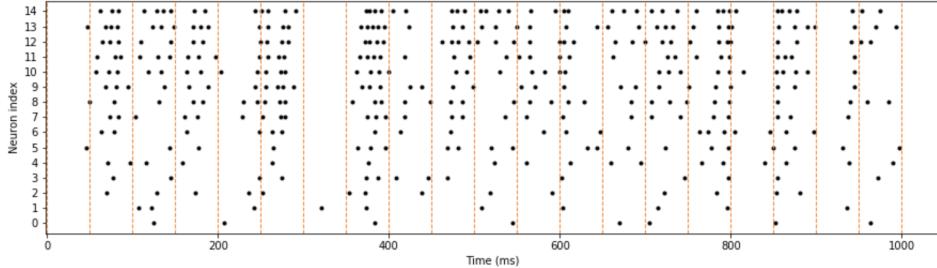


Figure 6: Raster plot of the 15-simplex network of regular spiking neurons, synapse weight 10, and time bin size 50ms

4 Experiments and Results

All experiments in this section are done with hyperparameters as in Table 3 from the Appendix. For similar experiments with different hyperparameters, please refer to the Appendix, or generate your own experiments using our code, available on github!

4.1 Entropy and number of parents

This first experiment examines the relation between the entropy of the distribution corresponding to a neuron N and the number of parents of this neuron, as well as the types of connections between the parents. We investigate whether one can estimate the number of parents of a neuron by simply knowing its entropy.

Four parent network types and varying sizes (number of nodes) are considered, in increasing complexity (density of connection): no connection, torus connection, simplex connection, full connection, as described in Section 3.2.2 and Figure 4. For a given parent network of a given size, we define another network consisting of the parent network, together with one node c with one incoming edge (p, c) for each parent node p in the parent network.

From this new network we follow the steps in Sections 3.2.2&3.2.3 to generate the spiking distribution of each network. In Figure 7 we plot the entropy (in bits) of the child neuron by the number of parents for the different parent connection types.

The most important determinant of the entropy of the child neuron in this graph seems to be the density of the parents connection. Too high density will produce chaotic behaviour and lead to a very low entropy as seen in the full and simplex connection case, whereas a lower density will lead to seemingly asymptotic entropy.

We first notice the degeneration of the case of fully connected parents. The other three connection types have a well behaved plot, but not the fully connected connection type. The fact that the entropy drops and stays low for high number of parents is simply due to the fact that the child neuron fires as many times

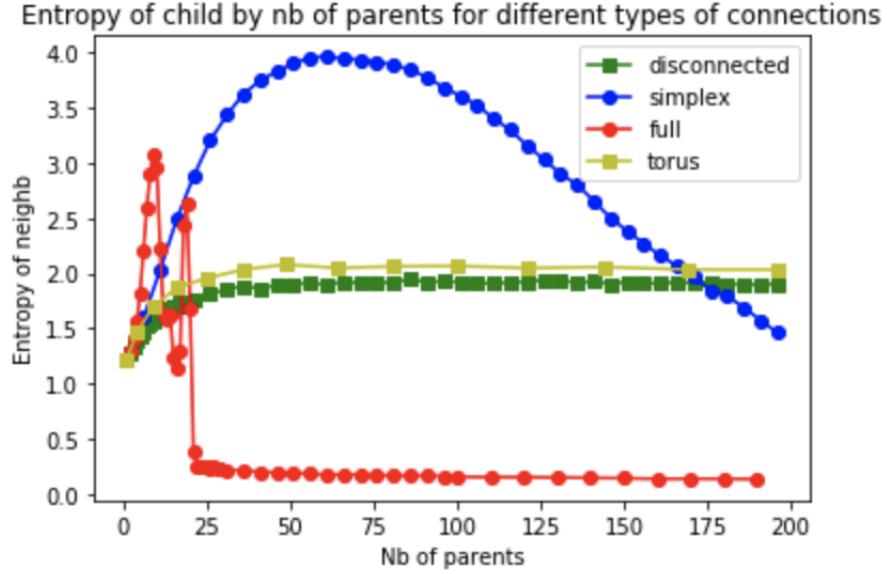


Figure 7: Entropy of child neuron by number of parents, for different parent types.

as it possibly can in a time bin, so the number of spikes for each time bin is constant and therefore the entropy is low.

The simplex case has asymptotic behaviour similar to the full connection, which is probably a result of a high density of connections. Indeed for a high number of parent the child will fire constantly. The peak value of the entropy is interesting. It seems that the high order of organization allows the entropy to reach a high value. This means that the signal produced by the neuron is very random: the neuron spikes in many different ways with a relatively high probability.

The disconnected and torus connection types behave quite similarly. In contrast to the other two connection types, the entropy does not seem to tend to zero as the number of parents increases. They both admit an asymptotic behaviour, the torus a little higher than the disconnected type, certainly due to its higher density of connection. It seems that low connection density between parent neurons makes the dynamics of the network stabilize. Note that this does not mean that the spike count of the child neuron is stable, but that its distribution is similar, indeed in each of the examples the expected spike count increased with the number of parents.

4.2 Mutual information and number of common parents

In [19], a study of the correlation between firing patterns of different neurons organized in simplex networks of different dimensions was made. The highest

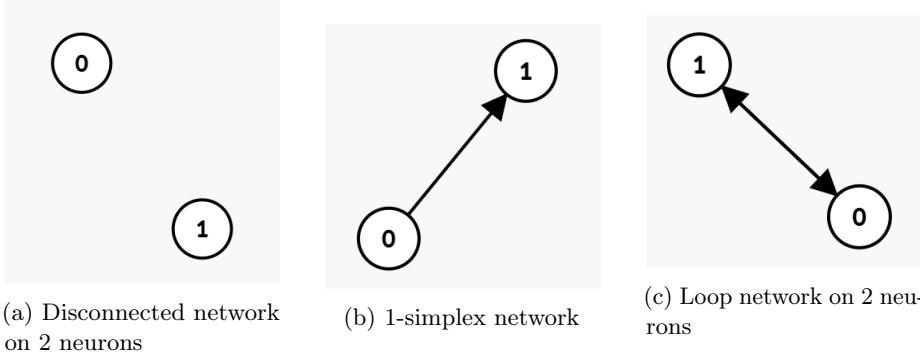


Figure 8: Different 2 neuron networks

correlation was found between the two last neurons for any simplex dimension. The study was reproduced in [17] using tools from information theory, where the simplex was stimulated at the source (neuron 0). Here we ask the question of how to maximize the mutual information between two given neurons, by changing the connection between these, the number of common parents, and the type of connection between their parents.

4.2.1 Mutual information of two neurons

In this section we consider three networks consisting of two neurons each, which can be seen in Figure 8. The first is two disconnected neurons, one is a network with a single directed connection, and the last one is a network with two connections forming a loop.

We compute the information-theoretic measures associated with these networks in Table 1. We notice that the mutual information increases with the number of connections. Indeed two disconnected neurons will have almost no mutual information, as expected; they should be independent. The mutual information is multiplied by 10 for one connection, and multiplied by 40 with a 'loop' connection. This indicates that mutual information can detect connection types between two neurons. Notice however that in all cases, the mutual information of each pair is only a small fraction of the entropy of each neuron: the maximum amount of shared information is only 3% of the entropy of a neuron. Therefore in relative terms, the connection types between two neurons does not have a significant impact on the mutual information between the measures. Indeed, we will see next that other factors, such as the number of common parents, and the connection types between the parents can have a more significant impact on the mutual information between two neurons.

Measure name	Disconnected	1-simplex	Loop
Entropy of neuron 0	1.177	1.179	1.269
Entropy of neuron 1	1.177	1.240	1.273
Joint entropy	2.352	2.408	2.501
Conditional entropy of neuron 0 by 1	1.275	1.168	1.228
Conditional entropy of neuron 1 by 0	1.175	1.229	1.232
Mutual information	0.001	0.012	0.041

Table 1: Information measures of two neuron spiking distribution.

4.2.2 Mutual information and common parent network

In this section we proceed similarly to Section 2.2.1, studying the impact of number of common parents and common parent connection type on the mutual information. We construct the networks as follows: define a parent network with a certain number of neurons and connection type between them, then add two neurons, and add a directed edge/synapse from each neuron in the network to each one of the new neurons. Note that the new neurons do not appear as pre-synaptic neurons, and therefore are not connected to one another. This is similar to the previous setup, but we now have two children instead of one.

We proceed as described in Section 3.2 to generate the joint distributions where the network is defined previously, for each parent connection type described in Section 3.2 and Figure 4. From the joint distribution we compute the mutual information between the two children neurons and plot this against the number of parents for each connection type to obtain Figure 9.

As for the entropy curve in Figure 7, one first notices the anomaly of the fully connected parent type. The dynamics do not behave similarly when scaling the number of parents, as shown by the isolated peak in mutual information for 20 parents. For the other connection types however, the distributions scale nicely with the number of parents, as shown by the smoothness of the curves in Figure 9. As for the entropy curve, the network standing out is the simplex connection type between the parents. Indeed, although each mutual information curve seems to monotonically increase with the number of parents, the increase in the simplex connection type is much steeper and reaches much greater values.

One interesting value to consider is the portion of entropy consisting of mutual information for each neuron. For the disconnected and torus types this value will peak for large networks at around 20%, which is significant, but this value peaks at over 60% at around 80 parents for the simplex connection type. This means that the signals produced by the children are much more similar for children neurons of simplices than for disconnected or sparsely connected neurons. One interpretation is the following: seeing the parent networks as different ways to stimulate the children neurons, the children encode the stimulus more similarly when stimulated by the same simplex (as the high mutual information suggests)

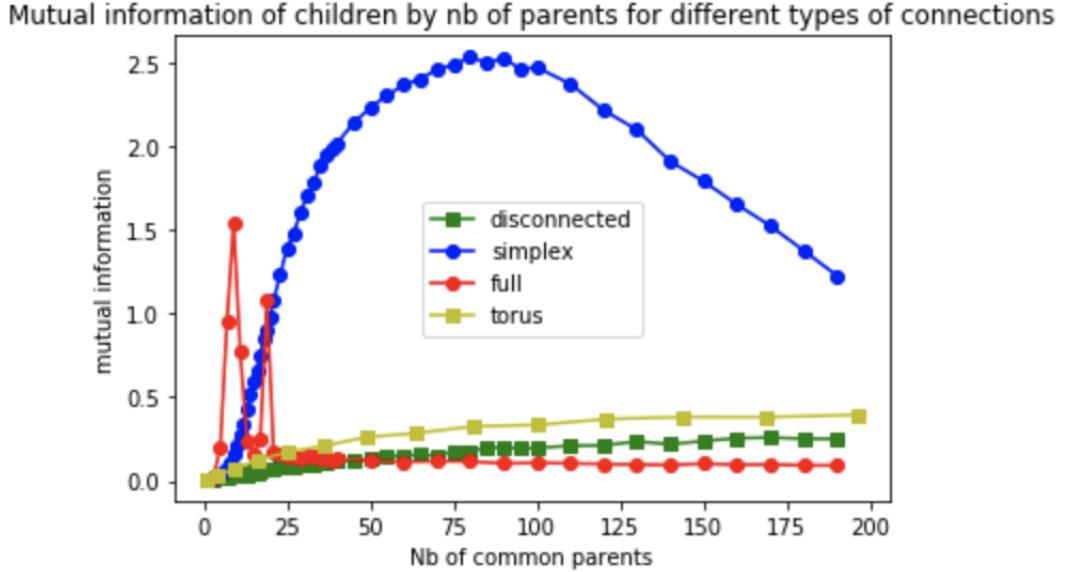


Figure 9: Mutual information for disjoint children by number of parents, for different parent connection types.

than they do in the case of other networks. One may wonder if this is a result of the high organization of simplices, or because of the density of the connections.

4.2.3 Adding connections between children

We changed the parent connection types in the previous section, so it is natural to look at the impact of connection between children. As in Section 4.2.1, we first add a single connection between the children, and then add two connections between them forming a ‘loop’. The results for each parent connection type can be seen in Figure 10.

Notice first that for each parent connection type, the plots for each children connection type have a similar shape. This suggests that children connections do not have a fundamental impact on the dynamics of networks. There is however a noticeable difference between sparsely connected networks (torus and disconnected) and densely connected networks (simplex and full). Indeed the latter seem less sensitive to changes in children connection types, as the network dynamics are already defined by the many connections between their parents. The former exhibits an interesting phenomenon already observed in Section 4.2.1: connections between two neurons increases their mutual information. One difference between disconnected and torus connection type is that the disconnected curves seem to stay distant from one another, whereas the torus connection type seem to converge after 125 parents. This is especially true for the loop connection in the disconnected case, which stays at 0.05-bits (16%, in relative

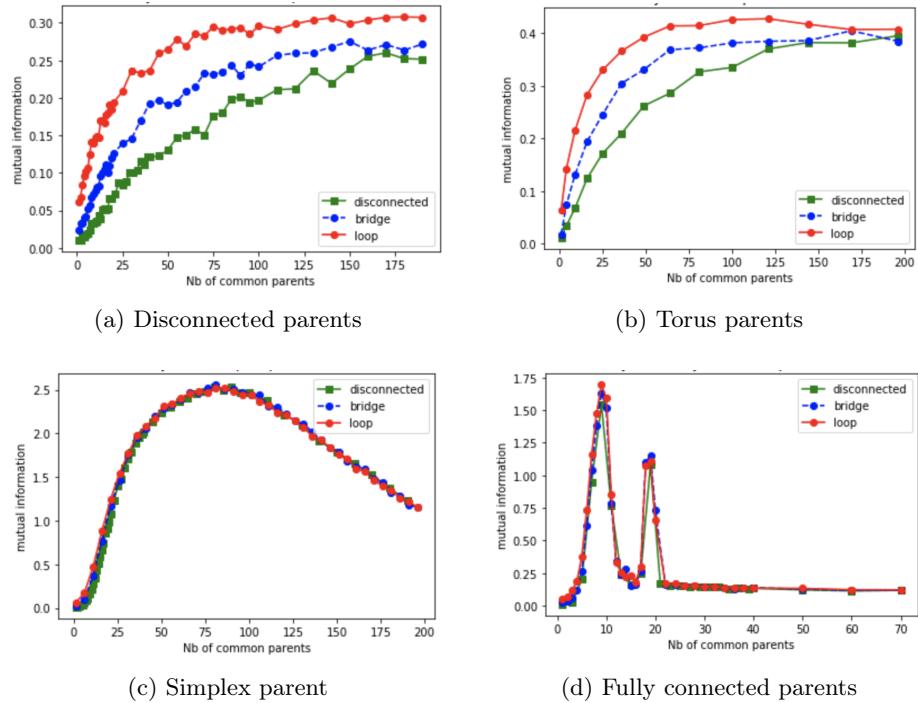


Figure 10: Mutual information of two children by number of parent, for different parent connection types and different children connection types

terms) above the other two connection types. One could again argue that this is due to the density of connections between neurons.

Note that the disconnected curves seem much less smooth than the torus curves, which is partly due to the fact that the torus curves are plotted with fewer points than the disconnected curves. For more details on how these graphs are generated please refer to the notebooks in the appendix!

4.3 Conditional entropy, conditional mutual information, and predictability of neurons.

Conditional versions of information measures have the advantage of being interpretable as mentionned in Section 2.2.1, and are therefore natural tools to use in studying interactions of several random variables. We have already used conditionning implicitly when computing mutual information of two children in Section 2.2.2. We now use it explicitly by analyzing the three-way interactions between the two children and their common parents. We finally see an example showing that one has to be cautious when conditioning on spiking data.

4.3.1 Conditional entropy and neuron predictability

To analyze the impact of parents on children, we first focus on a specific class of networks, namely the disconnected parents with two disconnected children for varying number of parents. We plot several information measures relating parents and child in Figure 11 as a function of the total number of parents in the network.

The blue curve in Figure 11 is the entropy of a child conditioned on its parents, measuring the randomness of the child's firing, given the state of the parents. Interestingly, the randomness starts by increasing from one to two parents and then decreases to an almost zero value at around 13 parents. The original increase is particularly surprising, as one would think that knowing more parents should decrease the randomness of the child firing signal. This means that the randomness gained by going from one to two parents is greater than the randomness lost by conditioning on two parents versus one. The later decrease is less surprising, as knowing the state of parents should indeed have an impact on the state of the child. It is interesting to note for around 13 parents, the conditional entropy is almost 0. This means that state of the child is almost deterministic in the state of the parents; we can guess with high precision the state of a child knowing the state of its 13 parents.

4.3.2 Bell shaped conditional mutual information

The red curve in Figure 11 is the mutual information of the two children neurons conditioned on the state of all parents. Notice that it is the difference between the blue and yellow curves. A point on this curve represents the additional

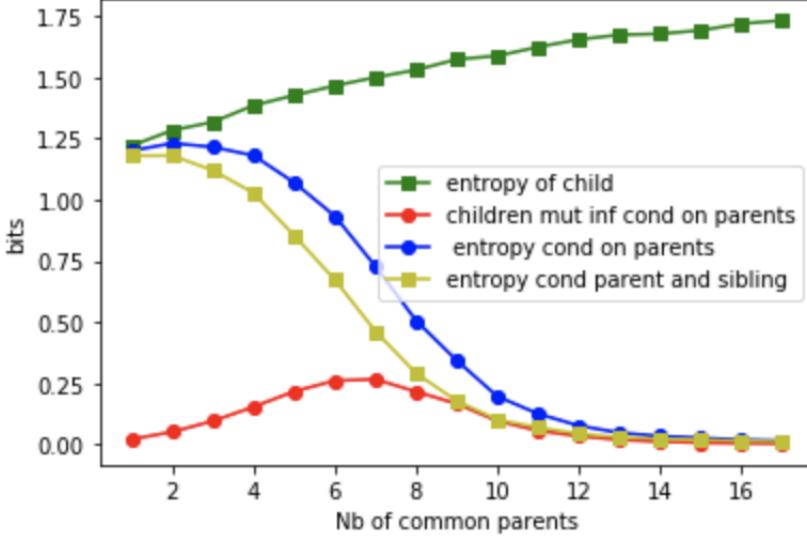


Figure 11: Entropy of child (green), conditional entropy of child by all parents (blue), conditional entropy of child by all parents and sibling (yellow), and conditional mutual information of siblings by parents (red). The x -axis is the total number of common parents.

information a child neuron provides about its sibling, knowing the state of all their common parents.

Firstly, the values are very high, much higher than the mutual information of two siblings for any network with less than 20 disconnected parents from Figure 9. This has a ‘synergistic’ interpretation using the co-information measure defined in 2.2.4 (this is inspired by Section 3.4 in [3]). We decompose the co-information of the triple (N_0, N_1, P) , where N_0 and N_1 are the children and P is the parent ‘vector’-random variable, in two ways in the following two decompositions of I_{co} :

$$\begin{aligned} I_{co}(N_0; N_1; P) &= I(N_0; N_1) - I(N_0; N_1|P) \\ &= I(N_0; P) + I(N_0; N_1) - I(N_0; N_1, P). \end{aligned}$$

The first line allows us to compute the co-information of the triple for each of the points on the curve considered, since the mutual information of the two neurons are close to 0 as seen in Figure 9, the quantity is almost equal to negative values of the red curve in Figure 11. The second line in the equation can be interpreted. The small negative (far from 0) values obtained from the first line tells us that $I(N_0; P) + I(N_0; N_1)$ is much smaller than $I(N_0; N_1, P)$, meaning that the sibling (N_1) and the parents (P) considered independently give much less information on the neuron (N_0) than if we consider them together. This gives rise to synergistic information being generated by the sibling together with the parents.

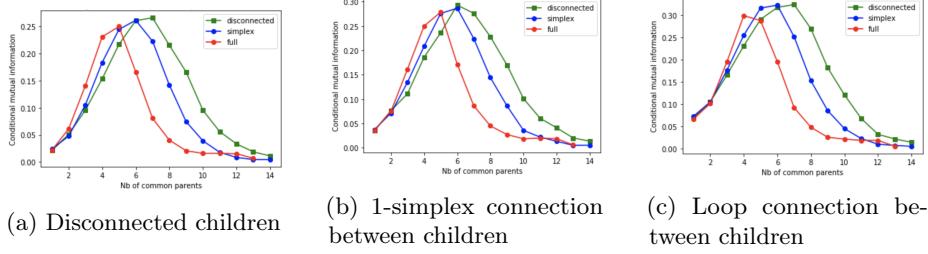


Figure 12: Mutual information between two neurons conditioned on all parents by number of parents, for different parent (red:full, blue:simplex, green:disconnected) and children structures (per plot).

Secondly, the bell shape of the red curve is surprising. The curve starts at around 0, reaches a maximum of more than 0.25 bits at around 7 common parents, and decreases back to 0. An interpretation of this is by revisiting the chain rule for mutual information:

$$I(N_0; P, N_1) = I(N_0; P) + I(N_0, N_1|P),$$

The conditional mutual information therefore measures the amount of information the pair P, N_1 provides us on N_0 that is not contained in P only.

This bell shaped curve is a general phenomenon we observed when observing how a group of neurons informs us on a specific neuron. We discuss this further in the next Section.

In Figure 12 we see how the structure of parents, and the structure of children impacts this curve. It seems that more activity (higher structural complexity) in the parents shifts the curve towards the left. Notice also that the peak value decreases slightly as the curve is shifted to the left. We also see how more structure in the children does not shift the curve, but its values increase.

In conclusion, the bell shape of the curve is very useful in analyzing the structure of networks. In the examples considered, the location of the peak informs us on the structure of the parents, and the values of the curve informs us on the number of common parents, and the structure between the children neurons.

4.3.3 Towards an information gradient

The x -axis in Figure 11 is the number of parents, and the conditioning is done on every parent. An increment in the x -axis can be thought of two distinct changes: in the network (number of parents) and in the conditioning (number of parents conditioned on). We decouple these two variables in Figure 13, where each curve corresponds to a network (labeled by the number of parents), and the x -axis corresponds to the number of parents conditioned on. Note that in the disconnected case we do not need to specify what parents are conditioned

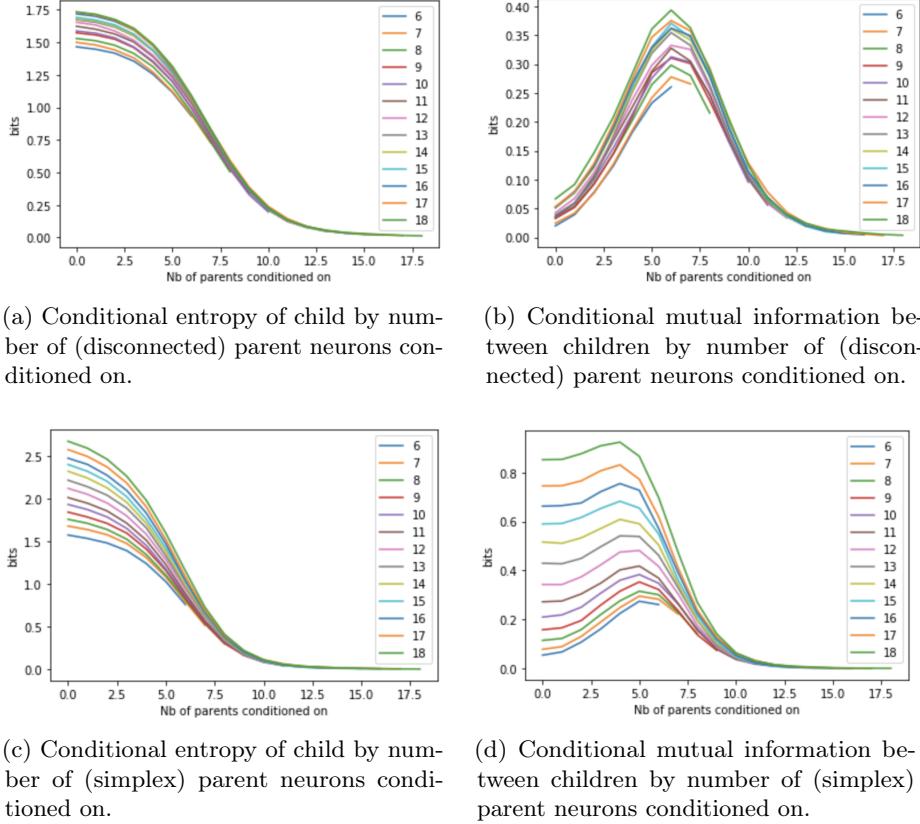


Figure 13: Conditional entropy and conditional mutual information, each curve corresponds to a network determined by the number of disconnected (a and b) or simplex parents (c and d).

on. In the simplex case, conditioning on n parents mean conditioning on the first n vertices of the simplex.

In general, the curves for the disconnected networks look a lot like in Figure 11. Indeed Figure 13a shows that knowing the state of 13 parents allows to guess the state of the children. We notice that the shape of the curve does not vary much depending on the colour, and their tails all converge. It seems that if the neuron has more than 13 parents, knowing the state of 13 (or more) of them seem provide enough information to guess the state the child is in. Figure 13c shows that in the case of simplex parents, only 10 parents are sufficient to make the conditional entropy close to zero.

As already observed in Figure 9, the mutual information between two neurons sharing fewer than 20 disconnected parents is small. It is when looking at neurons by groups that information can be obtained on a specific neuron. Indeed, the peak for each curve in Figure 13b is situated at 6 parents conditioned on, which

is close to the peak of the red curve in Figure 11. This phenomenon (the bell shape) is thus not a consequence of the number of parents, but a consequence of the number of parents conditioned on. The height of this curve does however depend on the number of common parents: the more parents in the network, the higher of the curve.

The shape of the curves Figure 13d distinguishes itself from the disconnected parent networks. The curves do not look exactly bell shaped, but the pattern of increasing till a peak value and decreasing to quasi zero values remains. The main difference is in the initial value, which is the mutual information of the two children neurons. This has already been seen to increase rapidly with the number of simplex parents. Doing the same analysis as in Section 4.3.2, we can interpret this as less important synergistic information being generated by the parents and one child about the other child. This could be a consequence of the fact that the neurons in the simplex share more information, and therefore generate less complementary information. The peak observed for the curves is located between 4 parents conditioned on for few parents and 6 for more parents. This happens much earlier than for the disconnected parent networks and shifts left as the number of parents increases, which confirms the assumption that it is a result of high parent activity.

4.3.4 Disconnected neurons are not independent

One subtlety with these measures is represented in Figure 14. We consider the disconnected network (no synapses) on 20 neurons and plot the conditional entropy and conditional mutual information of two neurons where we condition on a number of distinct neurons, represented on the x -axis. Note again that since the neurons are disconnected, the choice of neurons is not important, as long as the neurons conditioned on are distinct from the neurons previously considered.

We would like to first point out that the approach is slightly different to the previous approaches, as we are no longer focusing on the mutual information of two specific neurons (as we did with the children previously), but on the information a group of neurons provides about one specific neuron.

Notice that although the neurons have no relationship, we observe the same phenomenon as before. In particular, conditioning on enough neurons brings the entropy close to 0, relatively fast, and the conditional mutual information still has high values and a bell shape. This is surprising: a large enough group of neurons still can predict accurately the state of a neuron, and the bell shape of the red curve is still there. Similar conclusions as in Section 4.3.2 can be derived from these curves. This shows that in the model considered, disconnected neurons are not independent (in which case all curve should be flat).

This increases the significance on the nature of the information being captured. Indeed perhaps this is a result of the fact that each neurons considered have same hyper-parameters, namely the ones of regular spiking neurons, and perhaps the information captured is information given by these hyper-parameters. Another interpretation could lie in the fact that one can always reconstruct a random variable from many other random variables, as a function in several variables

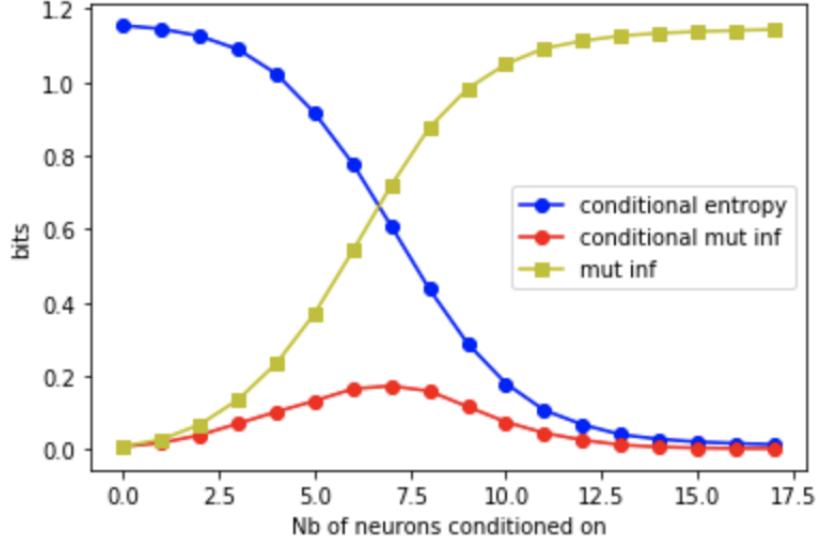


Figure 14: Conditional entropy of a neuron (blue), mutual information between two neurons (red) by number of neurons conditioned on. Mutual information between one neuron and a group of neuron by size of the group of neurons (yellow). All for a disconnected network on 20 neurons.

are more expressive than functions of few variables.

It is however important to note that the phenomenon differs for more structured networks in terms of the values and the shape of the curves, both in absolute terms and in relative terms. Indeed the conditional entropies reach an almost zero value much faster for simplex parents than disconnected parents.

Indeed the values for the blue curves seems to approach 0 more slowly (at 15 neurons conditioned on versus 13 before) than in more structure networks. Notice also that the peak of the red curve has value of 0.2 in the fully disconnected case whereas it is between 0.3 and 0.4 for more structured networks from Figure 13b. This is an increase from 17% to the total entropy of the neuron for the disconnected network to between 20% and 23% for the networks with disconnected parents and two children.

4.3.5 Integrating the bell shaped curve.

In this Section we related the bell shaped mutual information curve with the conditional entropy curve, through the intermediary of the mutual information of a neuron and a group of neuron.

We can explain the relationship among the blue, yellow and red curves in Figure 14 as consequences of information theoretic equations. Firstly, the yellow curve is obtained by subtracting the conditional entropy (blue) from the entropy of a neuron (first blue point).

Secondly, the chain rule shows that the yellow points are obtained by integrating (summing over points of) the red curve. The shape of the red curve is therefore extremely interesting, as it could give us an idea of the type of neuron providing more information about another.

We state the chain rule for mutual information between a neuron N and a group of neurons $G = \{N_0, N_1, \dots, N_k\}$:

$$I(N; G) = \sum_{i=0}^k I(N; N_i | N_0, \dots, N_{i-1}).$$

In light of this equation, the yellow points in Figure 14 are simply the sum of the previous red values. This means that the number of neurons needed to predict a neuron depends on the shape of this red curve. The wide and relatively flat red curve therefore has an interpretation that only little additional information is contained in each neuron. Whereas if the curve was thin and steep with a peak at 6 neurons could mean a higher concentration of information in the group of neurons selected. Note that the indexing of the neurons in G is essential in the decomposition of the mutual information. Changing the indexing gives a different red curve, but with the same area under it. We are very interested in this approach, to understand which neurons contain more information about a selected neuron, which we raise in Section 5.

4.4 Three Neuron Motifs

In this section we outline the notion of information simplex and different multivariate information-theoretic measures applied to the class of three-neuron motifs. This class of networks has attracted interest from the neuroscience community (ex: [15]). We construct the information simplex for each of these networks at rest. Since the networks are small, there is not a clear difference between the information measures of the motifs, but some subtleties can be observed. This is not surprising, since the information measures considered in this study tend to reflect the parent structure more than the sibling structure, as seen in Section 4.2.2 and Figure 10.

Figure 15a first distinguishes itself from the others by its particularly low values. One notices that the entropy of neurons 0 and 2 is lower than the entropy of neuron 1. Indeed, as seen previously, entropy of the neurons generally increases with input current. Not much can be read from the joint entropy of neurons alone, as the joint entropy of neurons 0 and 1 is between the joint entropy of the other two pairs. The connection is captured by the mutual information of neuron 0 and 1, which takes into account the high value of the entropy of neuron 1. Finally, the joint entropy of the three neurons is lower than for the other networks, reflecting the lower neural activity and is probably due to the number of different times a neuron can fire within a time-bin.

In contrast Figure 15b distinguishes itself with its large values, both for single neurons and for groups of neurons. The different mutual information values are

also high, and of higher order of magnitude than in Figure 15a, reflecting both the higher neural activity (from absolute values) and the higher synchronicity (from relative values, observed from the higher order of magnitude).

Unfortunately, it is hard to distinguish the simplex connection from the loop connection, although they are topologically distinct. Indeed, the three multivariate measurements have very close values, and one cannot use them to distinguish both networks. In fact, all mutual information values are also very close which makes it hard to distinguish. The only notable difference is in the joint entropies, which are higher in the loop network due to greater activity.

The stability of these measurements can be criticized since the symmetric networks, the full and the loop connections, do not have symmetric values. Indeed this can be seen in the third decimal. If one rounds to the second decimal, almost all the values do become symmetric as well.

5 Further Directions

5.1 More diverse Population of Neurons and Synapses

The main hyper parameters of the model that we have not changed are the biological ones. It would be interesting to change the neurons to ones with other spiking behaviour, and mixing different type of neurons. Raising the question on the interplay between of local dynamics of specific neurons and global dynamics of the whole network.

Similarly, a more biologically sound synaptic model would be a great use here. We believe that synapses do not cause instantaneous change in the membrane potential of the post-synaptic neuron, but has a cause a longer more diffuse reaction, could in fact exhibit interesting synergistic properties.

5.2 Other Measurements

An alternative to the generating the distribution by counting the number of spikes, would be to have one binary random variables per neuron indicating weather or not the given neuron spiked in that time bin. This more simple model will have an easier information theoretical interpretation, as the entropy of a neuron will be computable from the binary entropy function. these measures have been studied in [16]. It is unclear whether these random variables are expressive enough to record the structure of the underlying neural network. Similarly to [19], one could go even further and define a binary random variable for each group of neurons, indicating or not if all the neurons fired simultaneously. These random variables have a natural simplicial structure, which could encode much about the network.

5.3 Other networks

During this study, the high dimensional simplex stood out. This could be due to several properties of the simplex. Three of these are its high density, another

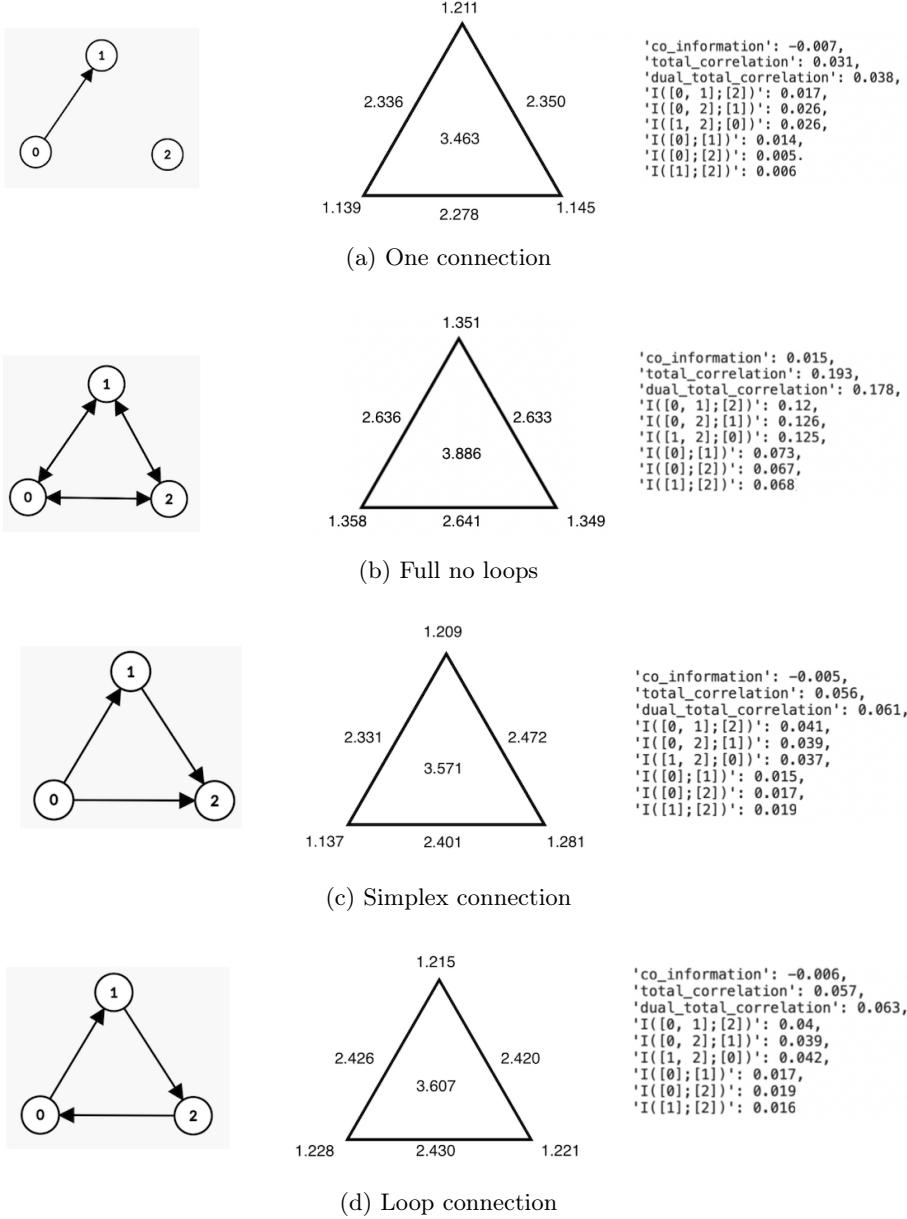


Figure 15: Network, information simplex and information theoretical measures of different three neuron motifs

could be the high dimensionality and a last one could be a simply connectedness property. The density of a network can be simply defined as the number of edges over the number of nodes, to test if equally dense networks exhibit similar properties one could look into random graphs with high probabilities to connect two nodes. The high dimensionality of the simplex is a purely structural property, the dimensionality of a network can be defined as the size of the largest simplex (clique) present in the network, to test this property one can consider several spaces obtained from glueing high dimensional simplices together. Another high dimensional space is the high dimensional torus, which is the n fold cross product of cycles. In this study we only considered the two dimensional torus, a product of two *cycles*. The high dimensional torus admits very complex topological features other than its dimensionality, this is due to the presence of cavities, this space is much more complicated than the high dimensional simplex from a structural property and is therefore very interesting to us.

5.4 Capacity and Structure of Channel

In Section 2.2.3 we introduced the channel capacity, which has the interpretation of how much information can be sent per channel use. The higher the capacity, the more efficient a channel is in sending information. An experiment we imagined is looking at how effective a neural network can be as a channel, meaning as a means to send a message from a source to a receiver.

We could design an experiment as having three groups of neurons, the source S , the channel N and the receiver R . Where every neuron in the source is connected towards every neuron in the channel, and every neuron in the channel is connected to every neuron in the receiver. Each of the three groups would consist of a network of several neurons and connections between them. The capacity of the channel would then be the mutual information between source and receiver $I(S, R)$.

The simplest case would be where the source, channel and receiver are disconnected networks. Another case would be a simplex channel and disconnected source and receiver. One could also vary the sizes of each of the three networks.

In this setting, several questions can be asked: given a channel N , what is the source/receiver pair that maximizes its capacity? Is it always the same source-receiver pair that maximizes capacity, or does it depend on the networks topology? Given a source and a receiver, what is the channel N giving the highest capacity? Is it always the same N that arises, or does it depend on the topology of the source and receiver?

5.5 Applications to network classification

A problem explored in [2] is network classification. Section 4.4 suggests that the information simplex is a valuable tool in distinguishing different networks. A limit to this approach is that we have used several hundreds of simulations to generate the distributions, but only one simulation is accessible when classifying a network. We would be interested in attempting to generate a distribution from

a single simulation. In order to make this possible, the simulation would have to be longer, and the time bin smaller, but constructing the information simplex of a given simulation could provide insight on the network's structure.

5.6 Information gradient and the high dimensional simplex

In Section 4.3.3 we raised the question of where information about a single neuron is stored. We saw that we can always predict the firing pattern of a regular spiking neuron by knowing 15 other regular spiking neuron, but in the case of the simplex, the 12 first parent neurons were enough to do so. It therefore seems that groups of neurons can give much information about a specific neuron. A first question would be to find the smallest possible group of neuron from which one can predict the firing of a single neuron. A second question would be to see how the information about the specific neuron is distributed among this group of neurons, which can be done by decomposing the mutual information in different ways using the chain rule as in Section 4.3.5. We are curious towards the role of the ordering of vertices in the simplex has to play in this distribution of information.

6 Conclusion

In this study we introduced several information theoretic concepts, we introduced a process to give a probabilistic model of spiking neuron data for given networks, and we applied the information theoretic tools to the probabilistic model, shedding light on the relation between structure and function of neural networks. The networks considered were organized as a group of parent neurons and a group of child neurons, where each parent had a connection to each of the children. Further structure was established between the parent neurons, as well as between the children neurons.

In Section 4.1, we saw that the entropy of the spiking behaviour of a single child neuron reflected the organization of the parents. Higher density and organization between parent neurons allowed for a wider range of possible behaviours of the child neuron, as witnessed by the high entropy values. In Section 4.2.2, we saw that although high density and organization between parent neurons provoked a more random behaviour in child neurons, the relationship between two disconnected children neurons was higher both in relative and absolute terms, as witnessed by high mutual information values. High dimensional simplices therefore have the property of communicating information reliably to several of their children. In Section 4.2.3, we saw that connections between children had a high impact on networks with sparsely connected parents.

In Section 4.3.1, we saw that one can use a number of neurons to predict the behaviour of a given neuron, as witnessed by a quasi-zero conditional entropy curve. In Section 4.3.2, we saw that a child can generate more information about another child when it is combined with a group of neuron, generating what we

called synergistic information, as witnessed by an analysis of the negative values of co-information. In Section 4.3.3, we saw that both the predictability of a child neuron behaviour and the synergistic behaviour observed were partly a consequence of the structure of the parent neurons, as witnessed by the shape of the conditional mutual information curves. In Section 4.3.4, we questioned the nature of a neuron’s predictability, by analyzing the behaviour of neurons that were not communicating, and concluded that the predictability of a neuron could be a result of the simplicity of our model that only considers regular spiking neurons. In Section 4.3.5, we provided an interpretation of the distribution of information of a specific neuron in a network, using the chain rule for mutual information to relate the predictability of a neuron to the synergistic behaviour in the information of this neuron.

In Section 4.4 applied our concept of information simplex to give an exhaustive information theoretical analysis of three neuron motifs.

Lastly, in Section 5 we outline further directions we would have liked to study but did not have the time to, including different ways to generate probabilistic models, different networks that seem interesting, and different information theoretical concepts that seem relevant to study spiking data.

[21]

References

- [1] S.-I. Amari and H. Nagaoka. *Methods of Information Geometry (Translations of Mathematical Monographs) (Translations of Mathematical Monographs)*. 04 2007.
- [2] J.-B. Bardin, G. Spreemann, and K. Hess. Topological exploration of artificial neuronal network dynamics. *Network Neuroscience*, 3(3):725–743, Jan 2019.
- [3] P. Baudot, M. Tapiá, D. Bennequin, and J.-M. Goaillard. Topological information data analysis. *Entropy*, 21(9):869, Sep 2019.
- [4] A. J. Bell. The co-information lattice. In *in Proc. 4th Int. Symp. Independent Component Analysis and Blind Source Separation*, pages 921–926, 2003.
- [5] F. R. K. Chung. *Spectral Graph Theory*. American Mathematical Society, 1997.
- [6] T. M. Cover and J. A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, USA, 2006.
- [7] C. Godsil and G. F. Royle. *Algebraic Graph Theory*. Number Book 207 in Graduate Texts in Mathematics. Springer, 2001.
- [8] D. Goodman and R. Brette. The brian simulator. *Frontiers in Neuroscience*, 3:26, 2009.
- [9] T. S. Han. Nonnegative entropy measures of multivariate symmetric correlations. *Information and Control*, 36(2):133 – 156, 1978.
- [10] A. Hatcher. *Algebraic topology*. Cambridge University Press, Cambridge, 2002.
- [11] S. Hoory, N. Linial, and A. Wigderson. Expander graphs and their applications. 43(4):439–561.
- [12] E. M. Izhikevich. Simple model of spiking neurons. *IEEE Transactions on Neural Networks*, 14(6):1569–1572, 2003.
- [13] E. M. Izhikevich. Dynamical systems in neuroscience: The geometry of excitability and bursting by eugene m. izhikevich. *SIAM Review*, 50:397–401, 01 2008.
- [14] R. G. James, C. J. Ellison, and J. P. Crutchfield. dit: a Python package for discrete information theory. *The Journal of Open Source Software*, 3(25):738, 2018.
- [15] P. Krauss, K. Prebeck, A. Schilling, and C. Metzner. Recurrence resonance” in three-neuron motifs. *Frontiers in Computational Neuroscience*, 13:64, 2019.
- [16] W. Li and Y. Li. Entropy, mutual information, and systematic measures of structured spiking neural networks, 2019.
- [17] Moalla, Skander. A multidisciplinary analysis of the link between structure and function in biological neural networks. Bachelor Thesis, EPFL Ecole Polytechnique, 2020. Accessible upon request.
- [18] E. Piasini and S. Panzeri. Information theory in neuroscience. *Entropy*, 21(1):62, 2019.
- [19] 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, 2017.
- [20] C. E. Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.*, 27(3):379–423, 1948.

- [21] N. M. Timme and C. Lapish. A tutorial for information theory in neuroscience. *eneuro*, 5(3), 2018.
- [22] S. Watanabe. Information theoretical analysis of multivariate correlation. *IBM Journal of Research and Development*, 4(1):66–82, 1960.
- [23] Wikipedia contributors. Mutual information — Wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=Mutual_information&oldid=997078708, 2020. [Online; accessed 3-January-2021].
- [24] Wikipedia contributors. Triangulation (topology) — Wikipedia, the free encyclopedia, 2020. [Online; accessed 4-January-2021].

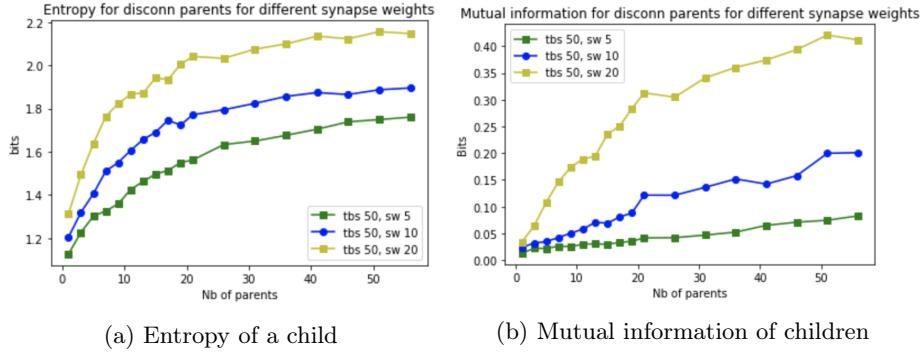


Figure 16: Entropy and mutual information for disconnected parents for different synapse weights

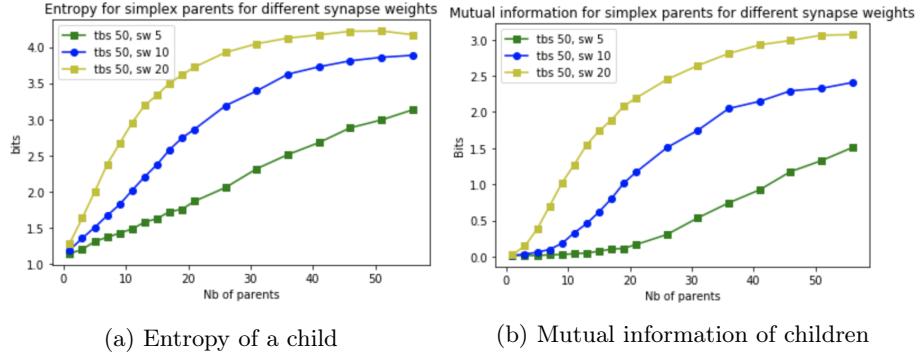


Figure 17: Entropy and mutual information for simplex parents for different synapse weights

7 Appendix

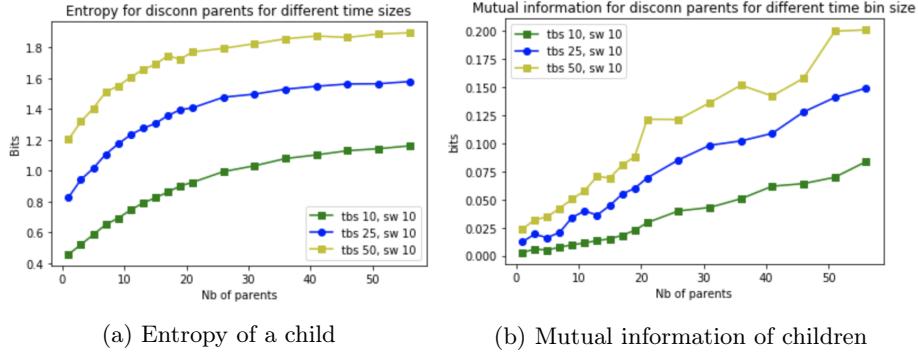


Figure 18: Entropy and mutual information for disconnected parents for different time bin sizes

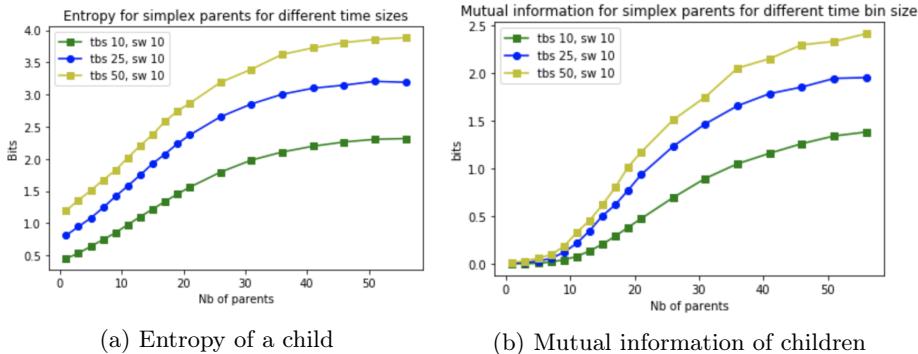
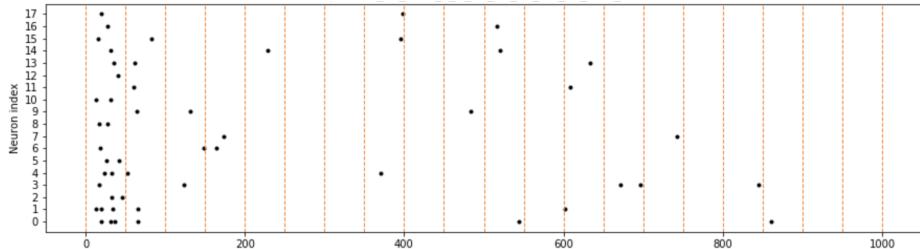


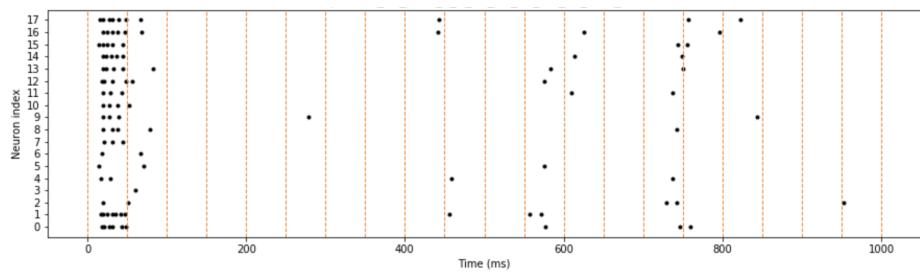
Figure 19: Entropy and mutual information for simplex parents for different time bin sizes

Parameter	Value
C	150
v_r	-75
v_t	-45
v_{peak}	35
k	1.2
a	0.01
b	5
c	-56
d	130

Table 2: Hyper-parameters for intrinsically bursting neurons.

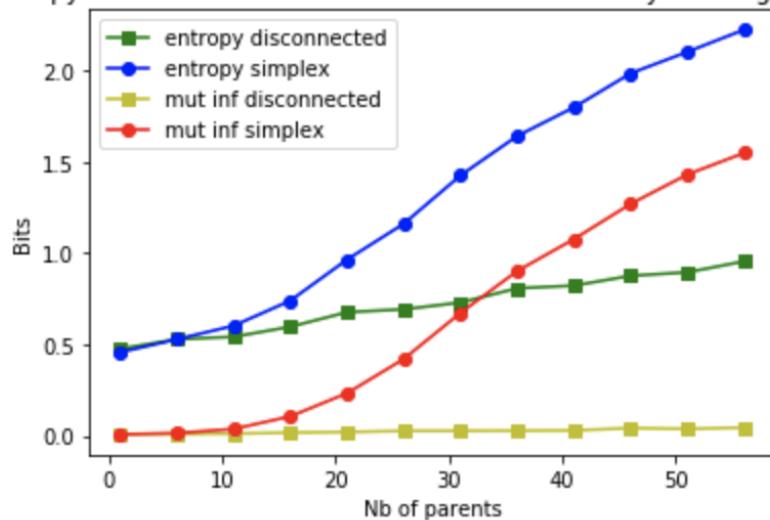


(a) Raster plot of 16 disconnected parents with two children indexed 0 and 1. All neurons are intrinsically bursting neurons



(b) Raster plot of 16 simplex parents with two children indexed 0 and 1. All neurons are intrinsically bursting neurons

Entropy of child and mutual information of intrinsically bursting neurons



(c) Entropy and mutual information for networks of intrinsically bursting neurons

Figure 20: Network, information simplex and information theoretical measures of different three neuron motifs

Parameter	Value
C	100
v_r	-60
v_t	-40
v_{peak}	35
k	0.7
a	-2
b	0.03
c	-50
d	100
Synapse weight	10
Time bin size	50ms
Duration	1000ms
Number of Simulations	500

Table 3: Hyper-parameter for networks of regular spiking neurons, used in Section 4.