PERSISTENT HOMLOGY ANALYSIS OF COGNITIVE MAPS

SYLVIA DURIAN

Abstract. This project is motivated by Dabaghian et al. [1], in which they use simulated data to test the hypothesis that cognitive maps encode topological information. In this paper, we apply persistent homology to experimental data from [4] in order to extend the study by [1] of the type of information encoded in cognitive maps. We find evidence that healthy mice create cognitive maps that are not dependent on geometric features and that become topologically accurate within a biologically reasonable time period. In contrast, there are intriguing hints that autistic mice may rely more heavily on geometric cues to build cognitive maps and may exhibit more variance in the time required to build topologically correct maps. However, there is not enough data for these results to reach statistical significance.

Contents

1. Introduction

In order to navigate through an environment, we must first have a fairly detailed internal map of the space, known as a cognitive map. Most neuroscience research on these cognitive maps assumes that they contain metrical information about the environment, such as the distance and angles between features of the environment, so that our brain can use the map to navigate in much the same way as we would

Date: SEPTEMBER 27, 2019.

use a street map. However, several considerations suggested to Dabaghian et al. [1] that cognitive maps may be constructed based on how different characteristics of the environment are connected to each other. Since connectivity and adjacency are topological relationships, they used simulated data to investigate whether the cognitive map encodes topological information. Their results suggest that cognitive maps may depend on both geometrical and topological features of the environment.

Here we extend their analysis from simulated data to experimental data from Talbot et al., [2] who collected this data to test whether autistic symptoms are caused by impaired neural coding. In particular, they asked if mice with autistic symptoms have more difficulty creating accurate cognitive maps than healthy mice. Although Talbot et al. analyzed their data based on the prevailing assumption that internal maps contain geometrical information, their data are also suited for our analysis. In this paper, we ask: does the topological accuracy of cognitive maps, and the time required to create those maps, differ in topologically equivalent but geometrically distinct environments? And does topological accuracy differ between healthy mice and mice with autistic symptoms?

To answer these questions, we will use persistent homology to interpret whether the brain captures topological characteristics. Persistent homology is an algebraic method of measuring topological features of spaces. Its importance lies in its ability to quantify the significance of different features, and so it is often used to de-noise and smooth data. Therefore, it is uniquely suited to our goal of analyzing internal maps from experimental data.

To apply persistent homology to our problem, the environment we are studying must be represented as a simpler space, called a simplicial complex. The idea of this is to build a shape out of points, line segments, triangles, and their n-dimensional analogs that has the same topological features as the environment.

With regard to organization, this paper begins in section 2 by giving some definitions in neuroscience which are relevant to the study of cognitive maps. In section 3, we develop simplicial homology and in section 4, we develop persistent homology on a simplicial complex. Section 5 modifies the Čech simplicial complex into a simplicial complex we can build using our data. Section 6 contains summaries of the relevant neuroscience papers [1, 2]. Section 7 outlines what we expect to see in our data analysis, i.e. computes the persistent homology of the actual environment for comparison against the persistence barcodes we will obtain from the data. Section 8 details the methods we use for data analysis, and section 9 contains the results of this analysis. Finally, section 10 discusses the implications of our results and suggests new directions.

2. Neuroscience Definitions

Without going into much detail about the neuroscience behind cognitive maps, here are some definitions from [5, 6] essential to our study.

Definition 2.1. When a neuron responds to a stimulus, we say that it *fires an action potential*. Formally, this means that the neuron sends an impulse along the membrane of a nerve cell, causing a change in its electric potential energy.

Definition 2.2. *Place cells* are neurons in the hippocampus that fire when an animal is at a specific location within its environment; this location is called the *place field*.

Since place fields are associated to place cells, place cells are believed to give the animal an internal representation of its environment, called a *cognitive map*. Place fields are generally ellipsoid in open field environments.

Definition 2.3. A *spike train* is a sequence of recorded times at which a neuron fires an action potential.

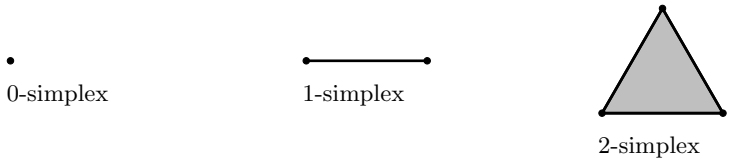
Since all action potentials are nearly identical in amplitude and width, neuroscientists generally use only the time of occurrence of the action potential to extract the content encoded by the neuron.

3. Simplicial Homology

Homology is a general way of associating algebraic groups to topological spaces which allows us to detect the holes (and higher dimensional analogs) in a space in an easily computable way. The simplest environment for homology is the simplicial complex. We give some definitions from [7, 8, 9, 10].

Definition 3.1. An *n-simplex* is a convex region in $\mathbb{R}^n$ spanned by $n+1$ points.

Intuitively, we can think of an n-simplex as the generalization of a tetrahedron to n-dimensions. Each simplex contains all of its lower dimensional simplices.



Definition 3.2. An *m-face* of an n-simplex A, where $m \leq n$, is a sub-simplex generated by $m+1$ vertices of A.

Definition 3.3. A *simplicial complex* K is a finite set of simplices satisfying:

1. If A is a simplex in K and α is a face of A, then $\alpha \in K$.
2. If simplices $A, B \in K$, then their intersection is either empty or a face of both A and B.

If L is a simplicial complex and every face of L belongs to K, then we call L a *subcomplex* of K.

Definition 3.4. Given an abelian group G and a set $(A_1^n, ..., A_k^n)$ of the n-simplices of a simplicial complex K, we define an *n-chain* c with coefficients in G as the formal sum

$$c = g_1 A_1^n + g_2 A_2^n + ... + g_k A_k^n,$$

where $g_i \in G$.

Remarks 3.5. We denote the group of n-chains by C_n.
Also, from now on we will assume that $G = \mathbb{Z}/2\mathbb{Z}$, which means, in particular, that we are not concerned with orientation.

Definition 3.6. Let A^n be an n-simplex in a complex K. The *boundary* of A^n is the $(n-1)$-chain of K over $\mathbb{Z}/2\mathbb{Z}$ given by

$$\partial(A^n) = A_0^{n-1} + A_1^{n-1} + ... + A_n^{n-1}$$

where $(A_0^{n-1}, ..., A_n^{n-1})$ are the $(n-1)$-faces of A^n.

Remark 3.7. We can extend the definition of boundary linearly to all of C_n. If $(A_1^n, ..., A_k^n)$ are the n-simplices of the complex K, then define

$$\partial_n = \sum_{i=1}^{k} g_i \partial(A_i^n)$$

Therefore, the boundary operator ∂_n is a homomorphism $\partial_n : C_n \to C_{n-1}$.

Definition 3.8. A *chain complex* is a sequence of chain groups

$$... \xrightarrow{\partial_{n+1}} C_n \xrightarrow{\partial_n} C_{n-1} \xrightarrow{\partial_{n-1}} ... \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0} 0$$

such that $\partial^2 = 0$.

Definition 3.9. The *n-dimensional homology group* H_n of the complex K over $\mathbb{Z}/2\mathbb{Z}$ is the quotient group $H_n = Ker(\partial_n)/Im(\partial_{n+1})$.

The elements of H_n are called *homology classes*.

Definition 3.10. The *n-th Betti number* β_n is the rank of the n-th homology group H_n.

Definition 3.11. A *triangulation* of a topological space X is a simplicial complex K together with a homeomorphism between K and X.

Theorem 3.12. *Homeomorphic spaces have isomorphic homology groups.*

See [8] for a proof.

Remark 3.13. This theorem, along with triangulation, makes simplicial homology really useful. It allows us to investigate homology in complex spaces by computing homology in the simple environment of a simplicial complex.

4. Persistent Homology

Given a simplicial complex, we may wish to know which of its topological features are the most important. Clearly, importance is a very subjective concept, but persistent homology gives a rigorous way to describe the relative significance of different features. We give some definitions from [11, 12, 13].

Definition 4.1. Given two simplicial complexes K and L, a *simplicial map* $f : K \to L$ is a map that sends simplices to simplices.

Remark 4.2. A simplicial map $f : K \to L$ determines, or *induces* a homomorphism on the homology groups $H_n(K) \to H_n(L)$ for each n. (See [8] for more details).

Definition 4.3. A *filtration* of a simplicial complex K is a nested sequence of subcomplexes of K that starts with the empty complex and ends with the complete complex:

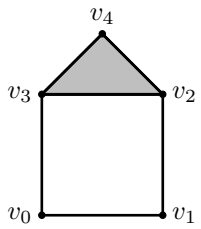
$$\emptyset = K_0 \subset K_1 \subset ... \subset K_m = K.$$

Definition 4.4. We say that a homology class α is *born* at K_i if it is not in the image of the map induced by the inclusion $K_{i-1} \subset K_i$. If α is born at K_i, then it *dies* entering K_j if the map induced by $K_{i-1} \subset K_{j-1}$ does not send α to 0 but the map induced by $K_{i-1} \subset K_j$ does send α to 0.

Definition 4.5. For α born at K_i and which dies entering K_j, the *persistence* of α is defined to be $j - i$.

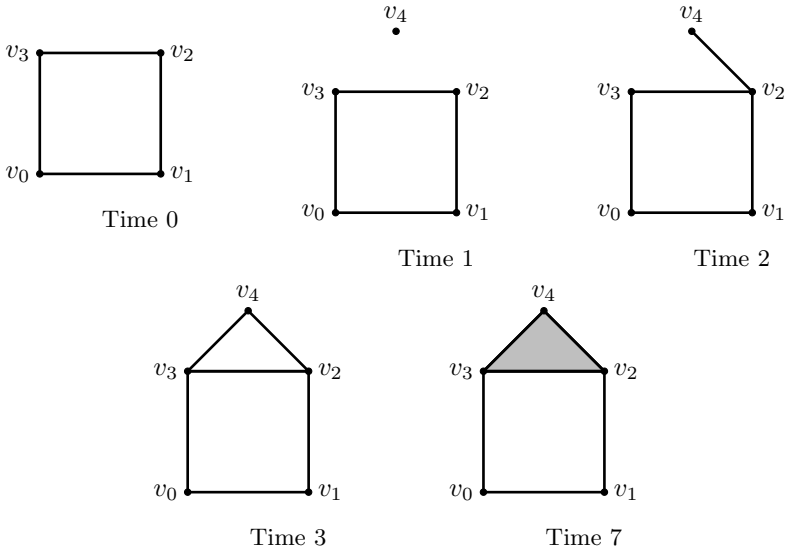
We encode this information in persistence barcodes, one for each dimension. Each bar code starting at i and ending at j represents a homology class that is born at i and dies at j.

Example 4.6. We borrow this example from [16] in order to illustrate how to interpret persistence barcodes.



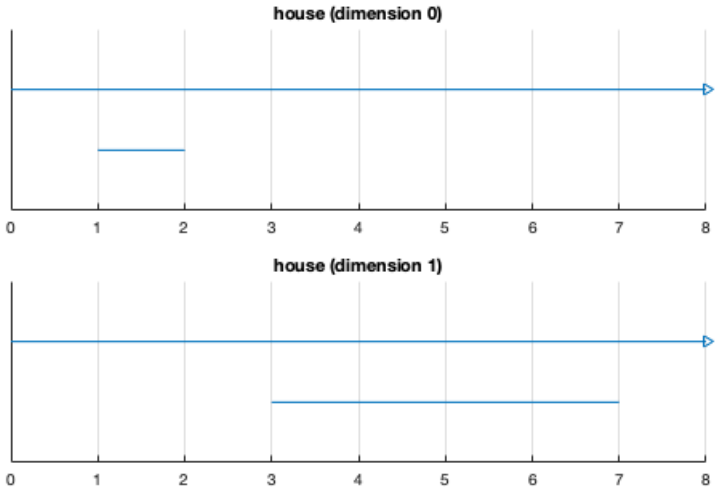
We want to compute the persistence of this "house" simplicial complex. The filtration we use will be dependent on time: at time 0, the vertices and edges of the square are built. At time 1, vertex v_4 appears; at time 2, the edge connecting vertices v_2 and v_4 is built; at time 3 the other roof edge appears. Finally, at time 7, the roof 2-simplex appears and the roof gets filled in.

FIGURE 2. Filtration over time



Using Javaplex [17], a package for computing persistent homology in Matlab, the persistence barcodes look like this:

FIGURE 4. Persistence barcodes for house simplicial complex



From the barcodes, we can reconstruct the topological features of the house: at time 0, a connected component and a 1-dimensional hole appear. A second connected component forms at time 1, but it joins to the first connected component at time 2. A second 1-dimensional hole forms at value 3, and closes at value 7. We say that there is one persistent 0-dimensional cycle and one persistent 1-dimensional cycle.

Note the connection between persistent bars and Betti numbers: it is an easy exercise to show $\beta_0 = \beta_1 = 1$. Since Betti numbers provide a rough numerical measurement of topological features, we can think of persistence barcodes as an analogue of Betti numbers in the sense that the barcodes tell us how many n-dimensional loops there are, and their length tells us their "importance."

5. Some Complexes

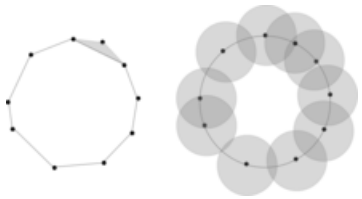
In this section, we modify the Čech simplicial complex into a simplicial complex that we can build using spike trains from place cells. Definitions are from [1, 14, 15].

Definition 5.1. Given a set of points $P = \{p_1, ..., p_m\} \subset \mathbb{R}^n$ and a real value $r > 0$, a simplex $A = \langle p_{i_0}, ..., p_{i_k} \rangle$ is in the *Čech simplicial complex N(P)* if and only if

$$\bigcap_{0 \leq j \leq k} \mathbb{B}(p_{i_j}, r) \neq \emptyset$$

where $\mathbb{B}(p_{i_j}, r)$ indicates the ball of radius r centered at p_{i_j}.

Remark 5.2. Čech simplicial complexes are especially useful in data analysis because they allow us to construct a simplicial complex on a topological space from a sampling of points in that space. For example, this is a Čech complex of a set of points sampled from a circle:



Definition 5.3. Given a finite collection of sets $\mathcal{F}$, its *nerve* is defined to be the simplicial set

$$Nrv\mathcal{F} = \{X \subset \mathcal{F} | \bigcap X \neq \emptyset\}.$$

The Čech complex $N(P)$ is the nerve of the collection of sets $\{\mathbb{B}(p_i, r) | p_i \in P\}$.

Theorem 5.4. *(Nerve Theorem). Let $\mathcal{F}$ be a finite collection of closed, convex sets in $\mathbb{R}^n$. Then the nerve of $\mathcal{F}$ and the union of the sets in $\mathcal{F}$ have the same homotopy type.*

See [14] for a proof.

Remark 5.5. In the context of constructing a simplicial complex to model hippocampal map formation, we can modify the definition of Čech complex slightly, based on the assumption that place fields are ellipsoid:

Given a set of place fields $\{PF_1, ..., PF_n\}$, a simplex $A = \langle i_1, ..., i_k \rangle$ is in N if and only if

$$\bigcap_{0<j\leq k} PF_{i_j} \neq \emptyset.$$

However, recall that the set of place fields is not actually accessible to us when we begin our analysis of hippocampal map formation; we only have spike trains from place cells. Therefore, we need to once again modify our complex. Since spike trains are time series, we want to construct our complex based on the notion of temporal overlaps between spike trains, rather than spatial overlaps between place fields.

Definition 5.6. Let $[0, T]$ be the time interval of data collection. Let $\{s_1, ..., s_N\}$ be the set of spike trains corresponding to N place cells, and fix $\epsilon > 0$ and $m \in \mathbb{N}$. We define the *temporal simplicial complex* τ by the rule that the simplex $\langle i_1, ..., i_k \rangle$ is in τ if and only if there exists a time $t \in [0, T]$ such that

$$\min_{j \in 1,...,k} |s_{i_j} \cap [t - \epsilon, t + \epsilon]| \geq m.$$

Intuitively, if the animal happens to visit a location that is covered by

$$\bigcap_{0<j\leq k} PF_{i_j} \neq \emptyset,$$

then there is a nonzero probability that k place cells will fire, or produce spikes, at roughly the same time. Because of how place fields are defined in relation to place cells, we can interpret any cofiring of the place cells as a spatial connection. For example, if at some time t two place cells cofire, then there is a link between the centers of the corresponding place fields (a 1-simplex is constructed), and so on as the number of place cells that cofire increases.

We assume that all instances of co-activity that occur in $[0, T]$ are "remembered" and can be used to establish the structure of the complex. This way, we obtain a natural filtration of the complex over time.

Remark 5.7. The Nerve Theorem (Theorem 5.4) still holds with our modifications. In other words, τ is homotopy equivalent to the union of all the place fields and so τ gives a complete topological description of the underlying space. Then, by Theorem 3.12, τ and the environment will have isomorphic homology groups. Therefore, analyzing the persistence on τ will give us the persistent topological features of the environment.

6. Background

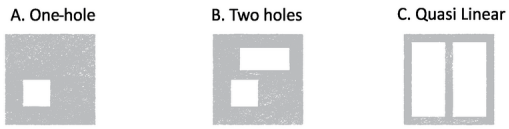
Here we summarize the neuroscience papers relevant to our project.

Although the reigning assumption in neuroscience has been that cognitive maps are geometrical maps which include distances and angles, Dabaghian et al. [1] hypothesized that the brain encodes a *topological* internal map. Recalling that as an animal explores a new environment, different place cells fire a series of action potentials in different areas of the environment, it seems that hippocampal output consists only of spiking patterns of many place cells and that other regions of the brain must somehow translate those patterns of cofiring into an accurate representation of the environment. More concretely, it is believed that the collection of place cells activated in a given environment produces enough place fields to cover the environment, thus creating a cognitive map. Moreover, because of the relationship between place cells and place fields, when two place cells fire at the same time it implies that the corresponding place fields are physically connected. This suggested to Dabaghian et al. that cognitive maps encoded by cofiring will be based on connectivity and adjacency which are topological relationships.

The researchers simulated spike trains of place cells by varying three characteristics of place cells: number of place cells, average firing rate, and place field size, and then simulating the random path of a rat through an environment. Based on past experiments, they thought that these variables would be most important in characterizing place cell activity. They investigated whether, and for which range of each of the three parameters, the simulated spike trains were able to form a temporal simplicial complex with the same topological features as the underlying environment.

They used three 2x2 m^2 environments:

Figure 5. Experimental scenarios



Note that environments B and C are homotopy equivalent, but geometrically distinct, allowing the experimenters to compare topological versus geometrical characteristics of the resulting simplicial complexes. The dense networks of gray lines in each scenario represent simulated trajectories, which actually nearly cover the environments. These trajectories were simulated to be random but non-preferential, so that there was no artificial circling or other favoring of one part of the environment over others.

Using the simulated spike trains, they built temporal simplicial complexes using the method in Definition 5.6. As previously noted, the resulting complex should be topologically equivalent to the environment by the Nerve Theorem. They tested whether this was actually the case using persistent homology: they expected that as the rat preliminarily explored the environment, there would literally be holes in its internal map, but continued exploration would allow the rat to discover the underlying structure of the environment and any extraneous holes would get filled in, or "die."

Dabaghian et al. found that spike trains corresponding to a small number of place cells (less than 150), with low firing rates (less than 10 Hz) and small place field sizes (less than 20 cm), produced complexes which consistently failed to capture the topological structure of the environment. Large place field sizes also impaired map formation, because they lowered the simulated rat's ability to identify where distinct parts of its environment began and ended. However, high numbers of place cells and high firing rates, along with medium sized place fields, reliably produced the correct topological features in 2-5 minutes. This timeframe, which we will describe as *map formation time*, is comparable to the time that rats and mice need to learn a simple environment of the same size [18, 19].

The researchers then constructed a 3 dimensional region using the range of the 3 parameters that led to an accurate topological map within a biologically reasonable time period, and called this a *learning region* L. Clearly the size of L depends on the complexity of the environment; in environment C, the holes are so big that the rat is forced to take a quasi-linear path around them. This makes it much faster for the rat to figure out the topological characteristics of the environment, and so the corresponding learning region is quite large. In contrast, the region L is the smallest for environment B, which reflects the fact that is more geometrically complex than C and more topologically complex than A.

Dabaghian et al. also found that L is stable– if one of parameters falls outside the range given in L but the other parameters are both in L, then map formation time changes minimally. Towards the boundary of the learning region, variations in map formation time increase drastically.

The stability and size of the learning region L confirm that this model was well-constructed. Firstly, the stability shows that the parameters chosen do indeed characterize place cell activity well. It also shows that the model is biologically reasonable since the dimensions of L match the biological ranges of the parameters. The fact that L is quite large is also in line with biological explanations of map formation– its large size suggests that it is possible for an animal to compensate for an unusual number of place cells, firing rate, or average place field size by upregulating the other parameters. From an evolutionary standpoint, this makes sense since the ability to form accurate cognitive maps is so essential to survival that it should not be able to be easily damaged.

These results show that the geometric complexity of an environment affects an animal's ability to build a topologically accurate cognitive map, suggesting that cognitive maps do depend on geometric features. However, their results also suggest that the brain encodes some topological information. In particular, the time it took to build topologically correct complexes in environments A, B, and C was both comparable to biological learning time and significantly shorter than the time it took for the simulated trajectories to cover the whole environment. This suggests

at least that cognitive maps are not formed by taking a simple random cover of the environment, as is generally assumed.

While these results are suggestive, it is essential to test these methods using experimental data instead of simulated data. An extensive data set was collected by Talbot et al. [2] for the purpose of comparing cognitive maps in healthy mice and mice with autistic symptoms. These data consist of spike train time series, just as in [1], which were collected from mice exploring several environments: a small and large box, a ring, and a circular open field that was either stationary or rotating. For more details on data collection, see [2, 3, 4].

To study autistic symptoms in mice, Talbot et al. worked with mice lacking the Fmr-1 gene (Fmr-1 null mice). Knocking out Fmr-1 causes the loss of a protein called fragile X mental retardation protein (FMRP), which results in intellectual disability and autistic symptoms. In particular, these symptoms are characterized by cognitive and behavioral inflexibility. Using these data, Talbot et al. [2] tested the hypothesis that neural coding is damaged by the absence of FMRP by comparing place cells in wild-type (healthy) and Fmr-1 null mice. They found that place fields produced by place cells in Fmr-1 null mice are hyper-stable and less spatially selective than those in healthy mice. In other words, it is much more difficult for Fmr-1 null mice to create accurate cognitive maps.

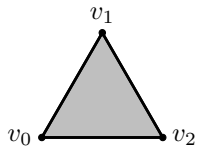
7. Expectations

Before we begin analyzing our data, it is worth outlining what we expect to see.

We will use data from the large box and circular open field (stationary). Although our environments are obviously 3-dimensional, we actually expect the cognitive maps to be 2-dimensional. This is because cognitive maps are created by a covering of place fields, and place fields form in the physical locations the mice has travelled. There are no 3-D bumps or depressions in the environments so the mice only travel in 2-D. We will therefore compare the complexes produced by the data to the 2-dimensional analogs of a cube and cylinder: a square and a disk. It is pretty clear that both are homotopy equivalent to a point, so we will only explain what we expect to see for the disk, and leave the square as an exercise.

Proposition 7.1. *The persistence barcodes for a disk (and a square) include one persistent cycle in 0-dimensions, and no persistent cycles in other dimensions.*

Proof. We triangulate the disk and use Theorem 3.12 to find its homology groups. The triangulation is just a 2-simplex:



To compute the homology groups, we first need to find the chain complex:

$$0 \xrightarrow{\partial_3} C_2 \xrightarrow{\partial_2} C_1 \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0} 0$$

We can see that there are three 0-simplices in this complex, so C_0 is isomorphic to $(\mathbb{Z}/2\mathbb{Z})^3$ with basis $\{(v_0), (v_1), (v_2)\}$. Similarly, there are three 1-simplices so C_1 is isomorphic to $(\mathbb{Z}/2\mathbb{Z})^3$ with basis $\{(v_0, v_1), (v_0, v_1), (v_1, v_2)\}$. There is one

2-simplex, so C_2 is isomorphic to $\mathbb{Z}/2\mathbb{Z}$ with basis $\{(v_0, v_1, v_2)\}$. There are no other simplices, so all other chain groups are 0.

It is immediate that $Ker(\partial_0)$= span$\{(v_0), (v_1), (v_2)\} \simeq (\mathbb{Z}/2\mathbb{Z})^3$.
Now, $\partial_1 : C_1 \to C_0$ is given by
$\partial_1(v_0, v_1) = (v_1) - (v_0)$
$\partial_1(v_0, v_2) = (v_2) - (v_0)$
$\partial_1(v_1, v_2) = (v_2) - (v_1)$
so $Im(\partial_1) = \text{span}\{(v_1) - (v_0), (v_2) - (v_0), (v_2) - (v_1)\}$. However, notice that

$$(v_2) - (v_1) = (v_2) - (v_0) - ((v_1) - (v_0))$$

so $Im(\partial_1) = \text{span}\{(v_1) - (v_0), (v_2) - (v_0))\} \simeq (\mathbb{Z}/2\mathbb{Z})^2$.
Then $H_0 = Ker(\partial_0)/Im(\partial_1) \simeq \mathbb{Z}/2\mathbb{Z}$.

Note that in $\mathbb{Z}/2\mathbb{Z}$,

$$\partial_1((v_0, v_1) + (v_0, v_2) + (v_1, v_2)) = (v_1) - (v_0) + (v_2) - (v_0) + (v_2) - (v_1) = 0$$

so $Ker(\partial_1) = \text{span } \{(v_0, v_1) + (v_0, v_2) + (v_1, v_2)\} \simeq \mathbb{Z}/2\mathbb{Z}$.
Next, $\partial_2 : C_2 \to C_1$ is given by
$\partial_2(v_0, v_1, v_2) = (v_0, v_1) + (v_0, v_2) + (v_1, v_2)$,
so $Im(\partial_2) = \text{span}\{(v_0, v_1) + (v_0, v_2) + (v_1, v_2)\} \simeq \mathbb{Z}/2\mathbb{Z}$.
Therefore, $H_1 = Ker(\partial_1)/Im(\partial_2) = 0$.

Finally, $\partial_2(v_0, v_1, v_2) = (v_0, v_1) + (v_0, v_2) + (v_1, v_2) \neq 0$, so $Ker(\partial_2) = 0$ and we have no 3-simplex, so $H_2 = Ker(\partial_2)/Im(\partial_3) = 0$.

Since all other chain groups are 0, all the Betti numbers are 0 except
$\beta_0 = \text{rank}(\mathbb{Z}/2\mathbb{Z}) = 1$.

In the persistence barcodes corresponding to this complex (and also to a disk by Theorem 3.12), we would therefore expect to have one persistent cycle in 0-dimensions, and no persistent cycles in other dimensions. We will only analyze persistence in 2-D since the cognitive maps are 2-dimensional, so in particular, we expect no 1-dimensional holes. □

Since [1] used simulated data and ignored several variables connected to place cell activity for the sake of simplicity of their model, we expect that our results will be qualitatively similar, but quantitatively different. Another reason for this expectation is that [1] used simulated rats, while the dataset [4] we will use has been collected from mice; little difference has been found in the basic properties of space representation between rats and mice, but rats are far superior at learning mazes [21].

8. Methods

We will use two environments: a square and a disk. While these are very simple shapes, they are geometrically distinct and homotopy equivalent, which allows us to test the hypothesis in [1]. As mentioned before, we will compute homology with $\mathbb{Z}/2\mathbb{Z}$ coefficients, which means, among other things, that we do not care about orientation.

Recall from Definition 5.6 the method we use to build temporal simplicial complexes (one complex per mouse per environment):

Let $[0, T]$ be the time interval of data collection. Let $\{s_1, ..., s_N\}$ be the set of spike trains corresponding to N place cells and fix $\epsilon > 0$ and $m \in \mathbb{N}$. The simplex

$\langle i_1, ..., i_k \rangle$ is in the temporal simplicial complex τ if and only if there exists a time $t \in [0, T]$ such that

$$\min_{j \in 1,...,k} |s_{i_j} \cap [t - \epsilon, t + \epsilon]| \geq m.$$

According to this definition, ϵ controls the time window for co-activity and m controls how many times a set of place cells must cofire in order for us to include the resulting simplex in the complex.

Assumptions.

(1) We choose $\epsilon = 1/8$ seconds because two or more cells are generally considered to be co-active if they fire within $1/4$ seconds of each other [20].
(2) We choose $m = 1$ because the data seems to be rather spread out.
(3) We assume the cognitive maps are 2-D because:
    - The maps are created by a covering of place fields.
    - Place fields form in the physical locations the mice have travelled.
    - The mice travel in 2-D.

We will analyze the data from [4] as follows: first, we take a random sample from the dataset so that we have 10 sets of spike trains each from healthy mice and from Fmr-1 null mice, 5 of which were taken in each environment. Then, we use the sets of spike trains to build temporal simplicial complexes, and analyze whether these complexes give the correct topological description of the environment using the persistent homology package in Matlab [16, 17]. Finally, we run hypothesis tests to investigate whether there is a significant difference in map formation between healthy and Fmr-1 null mice, and between environments.

9. Results

Here are the persistence barcodes from the temporal simplicial complexes constructed from the data. Each color represents a different mouse, and the horizontal axis is time in units of 100 microseconds.

FIGURE 6. Persistence barcodes from wild type (healthy) mice exploring square

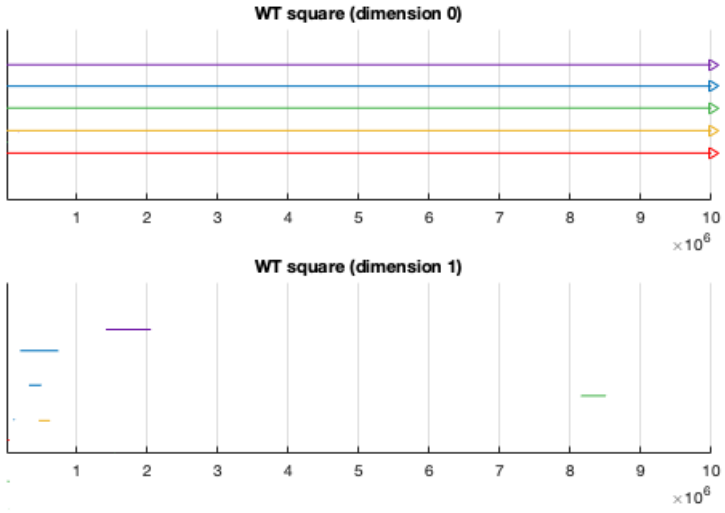


FIGURE 7. Persistence barcodes from wild type mice exploring disk

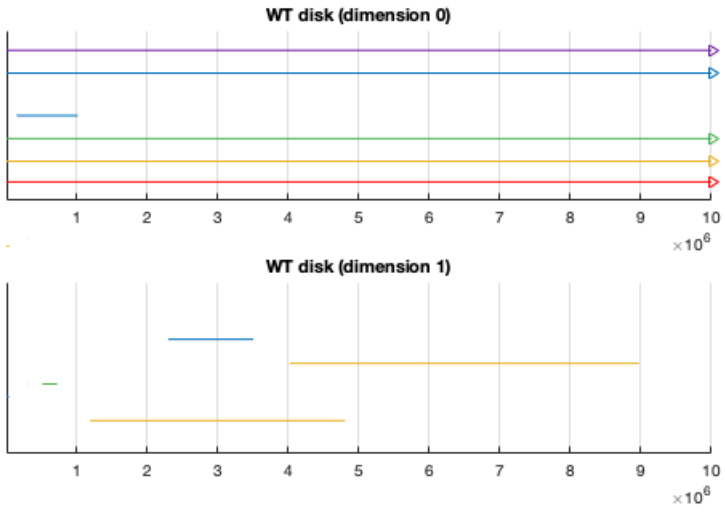


FIGURE 8. Persistence barcodes from Fmr-1 null mice exploring square

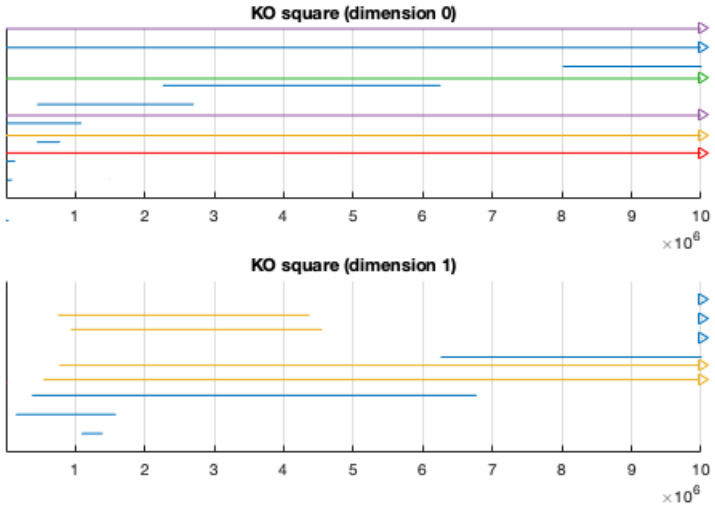
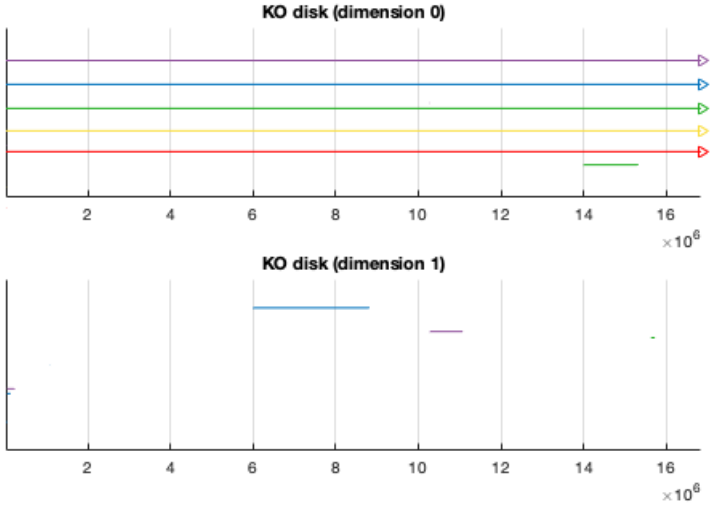


FIGURE 9. Persistence barcodes from Fmr-1 null mice exploring disk



As in [1], we use map formation time, or the time it takes to make a topologically accurate map, as an indicator of how well mice make the maps. We extract this from the persistence barcodes by taking the first time at which there is one persistent cycle in 0-dimensions and no other cycles. This information is better visualized using histograms. Again the x-axis is time in 100 microseconds.

FIGURE 10. Histogram of map formation times for wild type mice exploring square

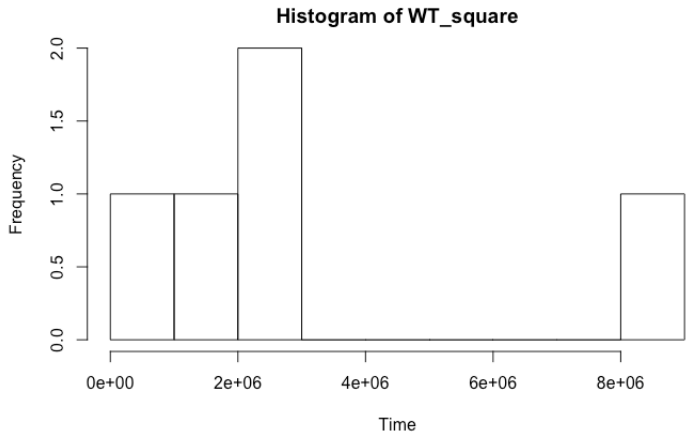
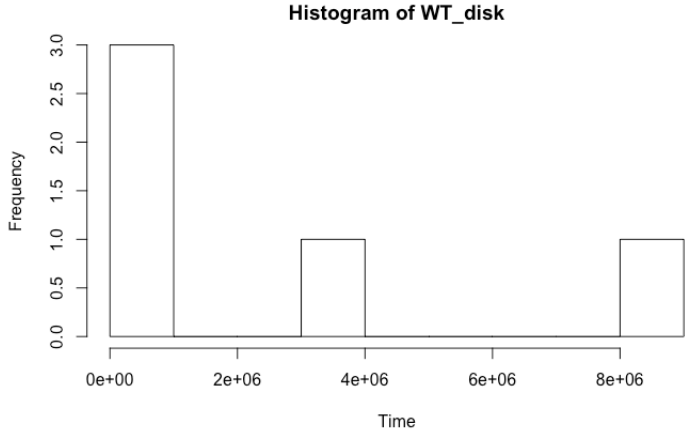


FIGURE 11. Histogram of map formation times for wild type mice exploring disk



From the persistence barcodes, we can see that three of the autistic mice did not make accurate cognitive maps during data collection time. We represent this in the histogram by using data collection time as "map formation time" for these mice, since collection time is certainly a lower bound on their map formation times.

FIGURE 12. Histogram of map formation times for Fmr1-null mice exploring square

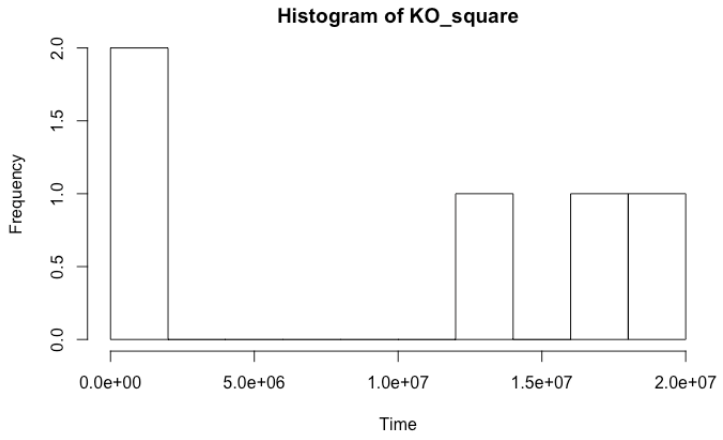
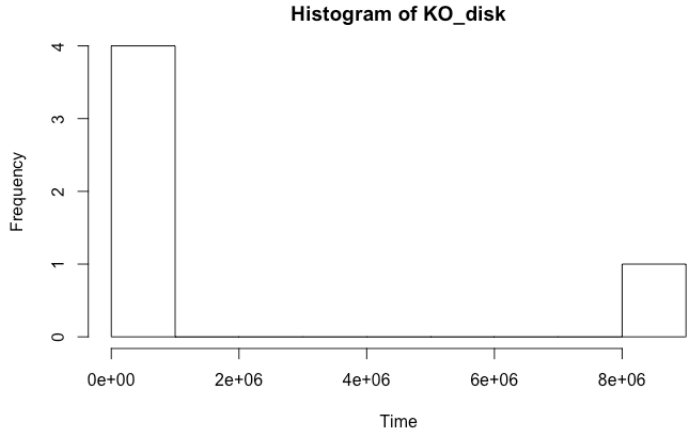
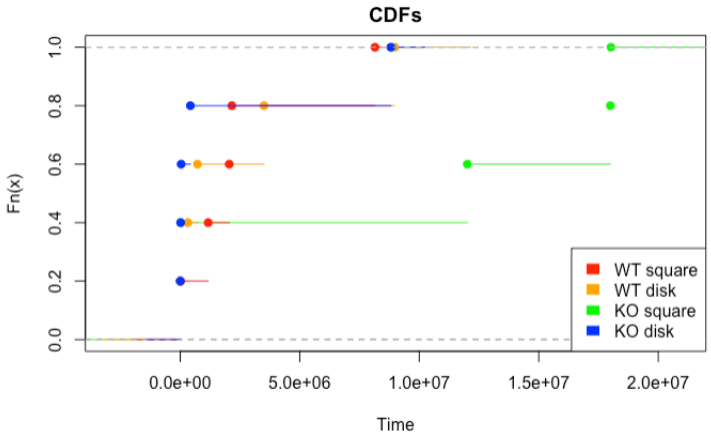


FIGURE 13. Histogram of map formation times for Fmr1-null mice exploring disk



Another way to visualize map formation times is by using cumulative distribution functions (CDFs). This kind of function returns the probability that a given distribution will take on a value less than or equal to the input. Therefore, CDFs are useful for comparing distributions. Here, we will use empirical cumulative distribution functions, which estimate the underlying CDFs given a sample of map formation times. The lines in the plot represent the range of values the true CDF could take.

FIGURE 14. CDFs



Inference

A formal way to test whether two distributions are the same is by using the Kolmogorov–Smirnov test. The Kolmogorov–Smirnov statistic quantifies a distance between the empirical distribution functions of two samples [22]. The null distribution of this statistic is calculated under the null hypothesis that the samples are drawn from the same distribution. From the statistic, we obtain a p-value that can be interpreted the same way as in a normal hypothesis test.

For any two combinations of the four scenarios, we use:
Null hypothesis: the distributions of map formation times in these two scenarios are equal.
Alternative hypothesis: the distributions are not equal.

9.1. **Comparison of environments using healthy mice.** Using the Kolmogorov–Smirnov test, the p-value is 0.873, so we fail to reject the null hypothesis and conclude that there is not a significant difference in the distributions of map formation times for healthy mice in the square and disk-shaped environments.

In other words, healthy mice are equally good at constructing accurate cognitive maps in square and disk-shaped environments. Since a square and disk are geometrically distinct but homotopy equivalent, this suggests that cognitive map formation does *not* depend on geometric features.

9.2. **Comparison of environments using Fmr-1 null mice.** Three out of five of the autistic mice sampled did not manage to make accurate cognitive maps by the end of data collection. As in the histogram, we use collection times as a lower bound on their map formation times so that we can perform a Kolmogorov–Smirnov test.

The p-value is 0.329, so we fail to reject the null hypothesis and conclude that there is not a significant difference in the distributions. While the difference is not statistically significant, the distributions in Fig. 12 (green and blue) certainly look dissimilar. This may point to something interesting going on – perhaps autistic

mice also use geometric cues to build cognitive maps– or the visual difference may be a result of our small sample size, as the p-value suggests.

9.3. **Comparison of healthy vs. Fmr-1 null mice using square environment.** Again, our Kolmogorov–Smirnov test uses collection times as a lower bound on map formation times in the three autistic mice who fail to build correct maps. The p-value for this test is 0.357 so we once more conclude that there is not a significant difference in the distributions for these scenarios. Although we cannot conclude anything interesting from the Kolmogorov-Smirnov test, the persistence barcodes (Figs 4, 6) suggest that autistic mice are more erratic in their map-making skills. All of the healthy mice form correct maps around the same time (except for one outlier), but three out of five of the autistic mice never make correct maps during data collection. Another data point that supports this view is that one of the autistic mice actually creates a correct map from time 0 (it only forms one 0-D cycle, which persists, and never forms any 1-D cycles)– it seems like this mouse is just guessing what the map should look like. However, we would need a much larger sample size and longer data collection periods in order to test this observation.

9.4. **Comparison of healthy vs. Fmr-1 null mice using disk environment.** For this Kolmogorov–Smirnov test, the p-value is 0.872 so we fail to reject the null hypothesis and conclude that there is not a significant difference in the distributions of map formation times for healthy and autistic mice in the disk environment. Interestingly, this p-value is much higher than the one generated for the distributions of healthy vs. Fmr-1 null mice in the square environment. Again, this may point to autistic animals using geometric features to build cognitive maps, but it could also be a result of our small sample size.

10. Discussion

Using persistent homology to analyze experimental data from Talbot et al. [4], we built on a study of cognitive maps by Dabaghian et al. [1]. Our results suggest that healthy mice do not depend on geometric features to create cognitive maps and that they are able to quickly build topologically correct maps. Not only are the distributions of map formation times in the two environments statistically indistinguishable, but the mere fact that all of the healthy mice make topologically correct maps lends support to the hypothesis in [1] that cognitive maps encode topological information.

Further, we tested the map-making skills of autistic mice and found differences, that, while not reaching statistical significance, are pronounced enough to merit gathering further data. Intuitively, it might make sense for autistic animals to have some kind of biological failsafe, such as using both topological and geometric information to build cognitive maps, which gives them a better chance of constructing correct maps.

In Section 7, we said that because we use methods from Dabaghian et al. [1], we expect our results to be qualitatively similar to theirs. However, our results are somewhat different, so here we will briefly discuss a possible reason for that difference. Recall that [1] used two square environments with different sized holes to test their hypothesis that cognitive maps encode topological and not geometrical features, and they found that the simulated rats built topologically accurate maps

much quicker for the environment with bigger holes than for the other one. Since the environments are homotopy equivalent, this would seem to be evidence against their hypothesis. However, the environment with bigger holes forced the rats to take a quasi-linear path around the holes so that they were able to cover this environment much quicker than the other one. This suggested to us that cognitive maps may depend on geometric *complexity*, but maybe not on other geometric features. In other words, if the environments had been such that the animal could move 2-dimensionally in both of them, perhaps the time it took to form topologically accurate cognitive maps would not have differed so dramatically. The environments we used, a square and a disk, both allowed for 2-D movement and there was no significant difference in map formation time, which supports this interpretation.

Another point of consideration is that the CDFs (Fig. 14) only capture map formation time, and not the rest of the information we can extract from the persistence barcodes. Although map formation time is a very important numeric quantifier of map-making ability, it could also be enlightening to investigate how the number of mistakes (dying cycles), and the time it takes to kill the dying cycles, differs between the environments and healthy vs. autistic mice. In other words, it would be interesting to look not just at when the mice figure out correct cognitive maps, but also *how*.

Although these preliminary results are encouraging, it is necessary to use a much larger sample with longer data collection times. In addition, it is crucial to test whether our results hold with more topologically complex environments, since we used such simple ones. This means using environments with 1-D holes if we continue with the assumption that cognitive maps are 2-D. This is one of our questions that needs to be addressed with further research.

Other questions include:

(1) Is there any intuitive reason (whether biological or mathematical) that healthy and autistic mice had very similar map formation times in the disk environment, but not in the square environment? Does this effect become more pronounced with larger sample sizes?
(2) Do these results hold with homology computed in coefficients other than $\mathbb{Z}/2\mathbb{Z}$?
(3) Are mice able to build cognitive maps that include 1-D holes embedded in a plane that they don't physically run over (e.g. windows)? Common sense says that they must be able to, but testing this would require developing a model of 3-D cognitive map formation that does not simply rely on the relationship between place cells and physical location of the mouse.
(4) How does the number of dying cycles, and their lengths, differ between environments and between healthy and autistic mice?
(5) If it is true that autistic mice use a different process from healthy mice to create cognitive maps, then it is worth analyzing the geometric accuracy of the maps they make. Is there a biologically and mathematically reasonable model to study this, i.e. can we make some kind of geometric analogue of the topological model used here?

This is a fascinating problem with many open questions which deserves further research.

Acknowledgments

It is a pleasure to thank my mentor, Foling Zou, for her help and guidance with this project. Many thanks also to Peter May for his insightful lectures in algebraic topology and for organizing this excellent REU.

References

[1] Y. Dabaghian, F. Memoli, L. Frank, G. Carlsson. A Topological Paradigm for Hippocampal Spatial Map Formation Using Persistent Homology. PLOS Computational Biology. 2012.
[2] Z. Talbot, F. Sparks, D. Dvorak, B. Curran, J. Alarcon, A. Fenton. Normal CA1 Place Fields but Discoordinated Network Discharge in a Fmr1-Null Mouse Model of Fragile X Syndrome. Neuron. 2018.
[3] Z. Talbot, F. Sparks, D. Dvorak, B. Curran, J. Alarcon, A. Fenton. Control of recollection by slow gamma dominating mid-frequency gamma in hippocampus CA1. PLOS Biology. 2018.
[4] Z. Talbot, F. Sparks, D. Dvorak, B. Curran, J. Alarcon, A. Fenton. Extracellular CA1 spike trains recorded from Fmr1-null and wild-type mice during free explorationin a variety of environments that include a small and a large box, a ring, and a circular open field that is either stationary or rotating. Talbot-Dataset.
[5] P. Dudchenko, E. Wood. Using Place Cells to Test Memory. Intro-to-Place-Cells.
[6] N. Tam. Neural Spike Train Analysis. Intro-to-Spike-Trains.
[7] J.P. May. A Concise Course in Algebraic Topology. University of Chicago Press. 1999.
[8] A. Hatcher. Algebraic Topology. Cambridge University Press. 2002.
[9] J.R. Munkres. Elements of Algebraic Topology. Perseus Publishing. 1984.
[10] H. Edelsbrunner, N.R. Shah Triangulating Topological Spaces. Triangulation-Survey.
[11] H. Edelsbrunner, J. Harer. Persistent Homology– a Survey. Persistent-Homology-Survey.
[12] H. Edelsbrunner, D. Morozov. Persistent Homology: Theory and Practice. Persistent-Homology-Survey-2.
[13] R. Ghrist. Barcodes: The Persistent Topology of Data. Persistence-Barcodes-Tutorial.
[14] C. Giusti. Applied Topology Lecture Notes. Simplicial-Approximations-Tutorial.
[15] Y. Wang. Topics in Computational Topology: An Algorithmic View. Commonly-Used-Complexes.
[16] H. Adams, A. Tausz Javaplex Tutorial. Javaplex-Tutorial.
[17] A. Tausz. Javaplex. Javaplex-Package.
[18] R. D'Hooge, P. De Deyn. Applications of the Morris water maze in the study of learning and memory. Brain Research Reviews. 2001.
[19] R. Morris, P. Garrud, J. Rawlins, J. O'Keefe. Place navigation impaired in rats with hippocampal lesions. Nature. 1982.
[20] G. Buzsaki. Theta rhythm of navigation: Link between path integration and landmark navigation, episodic and semantic memory. Wiley Online Library. 2005.
[21] V. Hok, B. Poucet, E. Duvelle, E. Save, F. Sargolini. Spacial cognition in mice and rats: similarities and differences in brain and behavior. Wiley Online Library. 2016.
[22] Statistics How To. Kolmogorov-Smirnov Goodness of Fit Test. KS-Test-Tutorial.