Detecting functional states of the rat brain with topological data analysis

Nianqiao Ju[1], Ismar Volić[2], Michael Wiest[3]

Abstract One of the cutting-edge methods for analyzing large sets of data involves looking at their "shape", namely their geometry and topology. In this paper, we apply topological analysis to data arising from a neuroscience experiment involving multichannel voltage measurements of brain activity in awake rats. Data points are viewed as a point cloud, with distance defined using channel correlations or a Euclidean metric. Exploratory data analysis reveals that the topological structure defined in terms of a Euclidean metric can distinguish between a coherent oscillatory brain state and the desynchronized awake state, by associating different Betti numbers to the different brain states.

Keywords: topological data analysis, mu rhythm, alpha rhythm, rat brain, persistent homology, Betti numbers, local field potentials, spike-and-wave

1. Introduction

Multi-channel neurophysiological recordings from the brain produce rich high-dimensional time series data from which neuroscientists attempt to distinguish different functional states and relate them to an animal or a person's behavioral capacities on the one hand and to underlying neural mechanisms on the other. Our goal is to explore whether topological data analysis, a new technique that has in recent years proved to be extremely fruitful in many fields, including in neuroscience (see [2] for a compilation of references), can reveal higher geometric structure in multichannel neural "local field potential" (LFP) voltage data and ultimately reveal information about functional states of the brain, or patterns of functional connectivity, that traditional methods cannot see. LFPs are analogous to electroencephalographic (EEG) recordings from the scalp, in that they reflect the electrical

[1] Nianqiao Ju, Department of Statistics, Harvard University, Cambridge, MA 02138, nju@g.harvard.edu

[2] Ismar Volić, Mathematics Department, Wellesley College, Wellesley, MA 02481, ivolic@wellesley.edu

[3] Michael Wiest, Neuroscience Program, Wellesley College, Wellesley, MA 02481, mwiest@wellesley.edu

activities of many neurons acting in concert, but they are "depth EEGs" recorded using electrode arrays surgically implanted into selected brain areas to better discern the sources of neurologically important "brain waves."

In this paper we focus on a test case comparing the topological structure of two known distinct states of the awake rat brain as measured by multisite LFP recordings. One is a state which can appear in immobile but awake rats, in which the LFP at multiple cortical and subcortical sites in the rat brain oscillates in a coherent high-amplitude rhythm with a frequency around 10 Hz [4,9,16]. This state has been referred to as "high voltage spike and wave discharges" [11,12,13,15] or informally as "mu rhythm" by analogy with a human brain rhythm in the same frequency range. For brevity in this study we will refer to this brain state as *mu*. We will compare episodes of this brain state to episodes of *non-mu* in which the brain is relatively "desynchronized," such that LFP fluctuations are smaller in amplitude and more broadband. Aside from being readily distinguishable in the LFP, these brain states have been shown to correspond to distinct modes of sensory processing [10].

The goal in this work is to apply topological analysis to the mu and non-mu data in hope that it can distinguish these states. This would support the possibility that topology might detect more subtle patterns that relate LFPs to behavioral and cognitive states.

Topology studies intrinsic geometric properties of objects, namely properties of the shape that remain unchanged after a continuous deformation. The most effective way of measuring and comparing such properties is to look at topological invariants of the space. A topological invariant is mathematical object, such as a polynomial or a group, that remains unchanged after the space is deformed. One of the most basic and effective class of invariants are homology groups. We will not define them precisely here since this is not needed for our purposes, but will say something about them in Section 2. For a precise definition, see [8] or [5]. Intuitively, homology groups keep track of the holes in a topological space. For example, the circle S^1 has a one-dimensional hole, while the sphere S^2 has a two-dimensional hole. Higher-dimensional topological objects might have higher-dimensional holes (in fact, the k-dimensional sphere S^k has a k-dimensional hole).

In topological data analysis, we view data as point clouds endowed with a certain geometry that in turn gives them the structure of a topological space. The points are intended to be thought of as finite samples taken from a geometric object, perhaps with noise. The geometry is provided by a distance function on the data, namely a notion of a distance between any two data points. The distance is defined using correlations between signals recorded from different parts of the rat's brain. From this distance function, one builds the topological space by means of a *Vietoris-Rips complex*. Finally, since we now have a topological space, we can compute its homology groups, thereby learning something about the shape of the data cloud from the information about its holes.

The paper is organized as follows: Some mathematical preliminaries, including basic background on homology and the Vietoris-Rips complex, are provided in Section 2. In Section 3 we describe the neurophysiological recording experiments

and data set. Results are presented in Section 4 and we summarize our conclusion in Section 5.

2. Mathematical background

Informally, a homology of a topological space X is the family of *homology groups*

$$H_0(X), H_1(X), H_2(X), \ldots \tag{2.1}$$

Each of them is a topological invariant that essentially counts the k-dimensional holes in X. The first homology group, $H_0(X)$, counts the number of connected components of the topological space, $H_1(X)$ counts the number of 1-dimensional holes, $H_2(X)$ counts the number of 2-dimensional holes, etc. For example, the homology groups of the circle S^1 are:

$$\begin{aligned} &H_n(S^1) = Z, \text{ for } n = 0, 1; \\ &H_n(S^1\} = \{0\}, \text{ for } n \geq 2. \end{aligned} \tag{2.2}$$

Here Z stands for the group of integers and {0} for the trivial group. More generally, for a k-dimensional sphere S^k we have:

$$\begin{aligned} &H_n(S^k) = Z, \text{ for } n = 0, k; \\ &H_n(S^k\} = \{0\}, \text{ for all other } n. \end{aligned} \tag{2.3}$$

What we mostly care about is the *rank*, namely the number of copies of Z, of each homology group, since this number essentially captures all the information about the group. The rank of the kth homology group is called the kth *Betti number*, denoted by

$$\beta_k = Rank(H_k(X)). \tag{2.4}$$

Thus β_k counts the number of kth dimensional holes. If $\beta_0(X)=1$, then X consists of a single connected component; if $\beta_1(X)=1$, then X has a single one-dimensional hole. A way to capture the number of holes is to see how many loops there are on the space that cannot be shrunk to a point (counting loops that can be deformed into one another as the same). An example that illustrates this is the torus $T^2=S^1\times S^1$, the Cartesian product of two circles (a hollow doughnut). It has one connected component, and so the 0th Betti number is 1; it has two holes because there are two essential loops (as shown in pink and red in the left panel of Figure 2) that

cannot be shrunk to points on the torus, so the 1-st Betti number is 2; and the space in the interior of T^2 is a two dimensional hole, so $\beta_2(T^2)=1$.

In order to make a topological space out of a data set, one first defines a notion of a distance on it. Namely, to any two points x_i and x_j in the data set, we associate a non-negative number $d(x_i, x_j)$ satisfying the usual properties of a distance function, i.e. of a metric. Then one endows the data set with the structure of a Vietoris-Rips complex, the standard way to make a topological space out of the metric in the context of topological data analysis. Briefly, the Vietoris-Rips complex of a data cloud X, attached to the parameter $\varepsilon > 0$, and denoted by $VR(X,\varepsilon)$, is the simplicial complex (a space built out of triangles, tetrahedra, and their generalizations) whose vertex set is X and where $\{x_i, x_2, \ldots, x_k\}$ spans a k-simplex if and only if $d(x_i, x_j) \leq \varepsilon$ for all $0 \leq i,j \leq k$. For an overview of the Vietoris-Rips complex and the idea of topological data analysis in general, see [1] or [3]. Figure 1 illustrates the Vietoris-Rips complex of a simple data cloud for various values of ε.

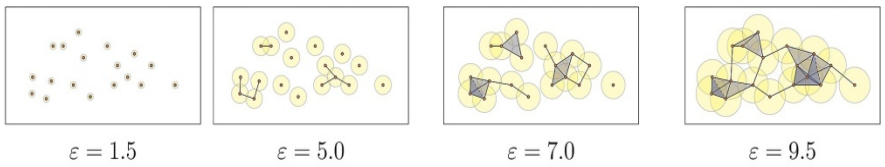

Fig.1. Example of Vietoris-Rips complexes at different ε (figure is taken from the Javaplex documentation). Connected components are constructed so that data points within ε of each other belong to the same component.

Once the data cloud has been given the structure of a topological space like this, we can compute its homology groups $H_k(X)$, $k \geq 0$. This can be done algorithmically through linear algebra using various online data analysis packages.

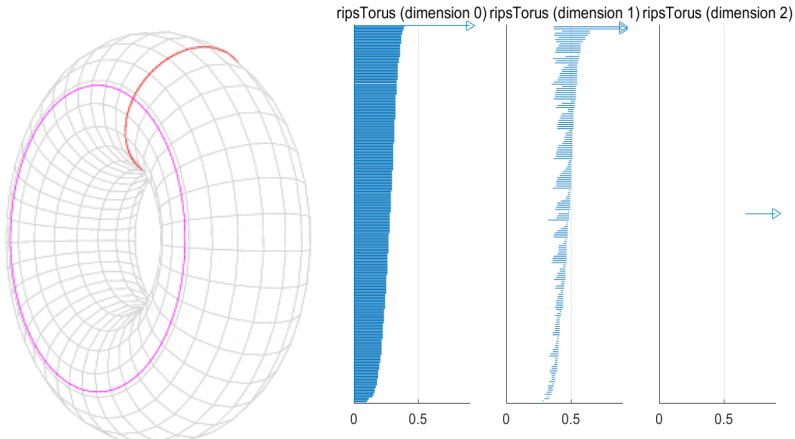

Fig. 2. A torus T^2 with $\beta_0(T^2) = 1$, $\beta_1(T^2) = 2$ and $\beta_2(T^2) = 1$. We see that the barcode plot shows exactly these the Betti numbers. To read the Betti numbers, we count the number of arrows in the barcode plots associated with each dimension.

The one used here was Javaplex [14]. Javaplex produces a *persistence barcode* for each homology group, with the number of bars that "survive" being the Betti number for that homology group. Figure 2 gives an example of the persistence barcodes for the torus. The interpretation is that the long bars are holes in the data cloud that appear for various values of ε, i.e. they are persistent, and this means that those holes are essential to the data cloud.

Note that all that is necessary to perform topological data analysis on a data cloud is the metric, i.e. the distance function; the rest is essentially automatically done by a computational tool such as Javaplex.

3. Materials and methods

Local field potentials (LFPs) were recorded at 16 parietal and 16 frontal sites in the cortex of a male Long-Evans rat while the rat passively listened to 100 ms duration tones of two different pitches, presented with equal probability in random order. The sample rate was 1000 Hz. Trials were defined as segments of LFP from 0.5 s before each tone until 1.5 s after the tone. For the present study to avoid confounds due to the two pitches we only analyzed trials in which the lower pitched tone (1500 Hz) was presented to the rat. We first rejected artifact trials automatically using a 1.5 mV threshold.

During the passive recording session the rat spontaneously went in and out of the synchronized ~10 Hz oscillatory state we are referring to as a mu-rhythm. Our goal is to compare the topology of mu and non-mu trials to see whether it can capture the difference in brain states. To identify mu and non-mu trials for the purposes of this comparison, one of us (MCW) with experience studying this brain state selected 126 *mu* trials and 136 *non-mu* trials based on visual inspection of one frontal LFP channel. The selected mu trials exhibited characteristic "spike-and-wave" patterns for the whole 2-second trial. Conversely, the trials selected as representative non-mu trials were free of the spike-and-wave oscillation for the entire trial.This procedure resulted in a set of 126 mu trials and 136 non-mu trials. Four examples of LFP recordings in each state are shown in Figure 3. We chose these two brain states as a test case for our topological analysis because they are clearly distinct in the LFP, even to an untrained eye.

Thus the total data set comprised a

$27 \times 126 \times 2001$ = (number of LFP channels) × (number of trials) × (number of time points)

3-dimensional grid for the mu trials plus a 27 x 136 x 2001 grid for the non-mu trial data. Further details about electrode implantation, recording coordinates, preprocessing, and other experimental procedures may be found in [6]. All procedures involving rats were approved by the Wellesley College Institutional Animal Care and Use Committee.

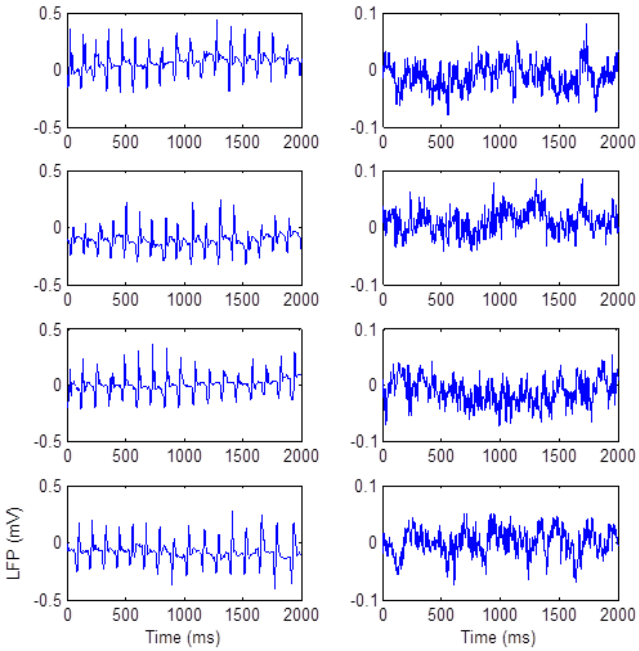

Fig. 3. The *left column* shows four examples of one local field potential (LFP) channel recorded from frontal cortex of an awake rat during episodes of an oscillatory brain state we refer to as mu. The *right column* shows four example trials recorded in the same rat during the same session, but while the rat's brain was in a relatively desynchronized state we refer to as non-mu. In every trial a brief tone stimulus was presented to the rat at 500 ms.

As a possible way to learn about functional connectivity between various parts of the brain, we analyze the data using a persistent brain network homology. For each trial, denote the data set as a string $C=(c_1, c_2, \ldots, c_n)$ consisting of n nodes where n is number of channels and each c_i is a 2001-dimensional vector whose coordinates are the LFPs at each ms of a 2.0 s trial. Inspired by an earlier paper [7], we calculate the distance matrix D based on correlation between channels, defined as

$$D_{ij} = \sqrt{1 - corr(c_i, c_j)} \tag{3.1}$$

where

$$\bar{c}_i = \frac{1}{2001}\sum_{t=1}^{2001} c_{it} \quad \text{and} \quad corr(c_i, c_j) = \frac{\sum_{t=1}^{2001}(c_{i,t} - \bar{c}_i)(c_{j,t} - \bar{c}_j)}{\sqrt{\sum_{t=1}^{2001}(c_{it} - \bar{c}_j)^2 \sum_{t=1}^{2001}(c_{jt} - \bar{c}_j)^2}} \tag{3.2}$$

is the sample correlation between signals from the *i* th and *j* th channel. The correlation, which is a number between -1 and 1, captures the linear relationship between the channels. If the correlation is close to 1, this would indicate the two channels are positively linearly related and "functionally connected." Figure 4 gives an example of a distance matrix for a sample trial.

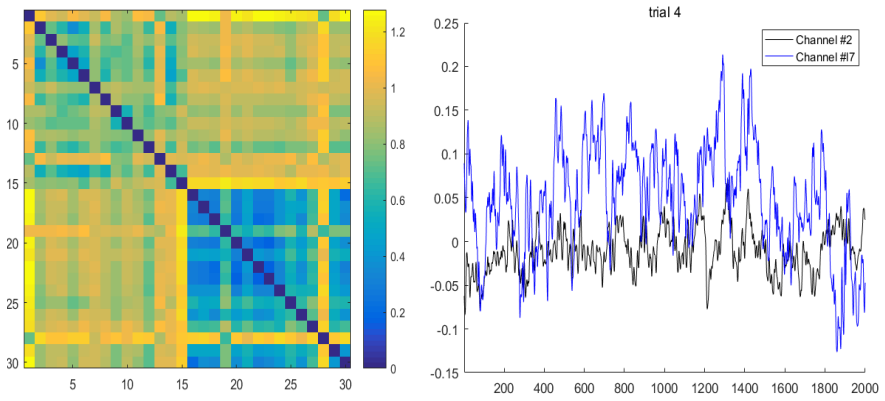

Fig. 4. *Left panel*: The distance matrix D for trial No.4 - a single trial in a session where the rat sat passively while listening to 2 different beeps played in random order. Channels 1-15 are frontal channels, and channels 16-30 are parietal channels. *Right panel*: Signals from two channels in trial No. 4, a frontal channel #2 and a parietal channel #7. The horizontal axis shows time in milliseconds and the vertical axis shows the LFP voltage in millivolts.

With the metric now defined, we can associate the topological space $VR(C,\varepsilon)$ to our data, and then compute its homology using Javaplex.

In addition to the metric described above, we also implemented the naive Euclidean metric, treating each trial as 2001 points collected from a 27 dimensional space, endowed with the standard Euclidean distance.

4. Results

In order to test the potential of topological data analysis for understanding multi-channel LFP neural data, we compared the topology of mu trials, exhibiting a high-amplitude rhythmic 10 Hz oscillation, to the topology of relatively desynchronized non-mu trials. Examples of the two LFP states are shown in Figure 3.

We take Trial 4, whose distance matrix and channels #2 and #7 are shown in Figure 4, as an example to illustrate our correlation-based topological analysis. We obtained $\beta_0 = 2$ and $\beta_1 = 1$ as the only nontrivial Betti numbers for this trial. Topologically, this means that the data has two connected components and that one of the components has a 1-dimensional hole, or an essential circle that cannot be shrunk within the data cloud. Because the distance we defined arises from

channel correlations, we believe the two connected components correspond to the two brain areas - the frontal and the parietal area.

We first used the correlation distance to analyze all 262 trials, and examine the resulting β_0 from the two groups. Unfortunately this metric turned out to be not revealing in distinguishing between mu and non-mu trials. We ran a Wilcoxon rank-sum test on the β_0's from the 262 trials to test the hypothesis that the two populations has the same distribution. This nonparametric test has a p-value of 0.0025, which means we can reject the null hypothesis at the 95% confidence level. We also ran a Student-t test (dof = 261) comparing the mean β_0 in each group. It returned a p-value of 0.001, supporting that the means are significantly different.

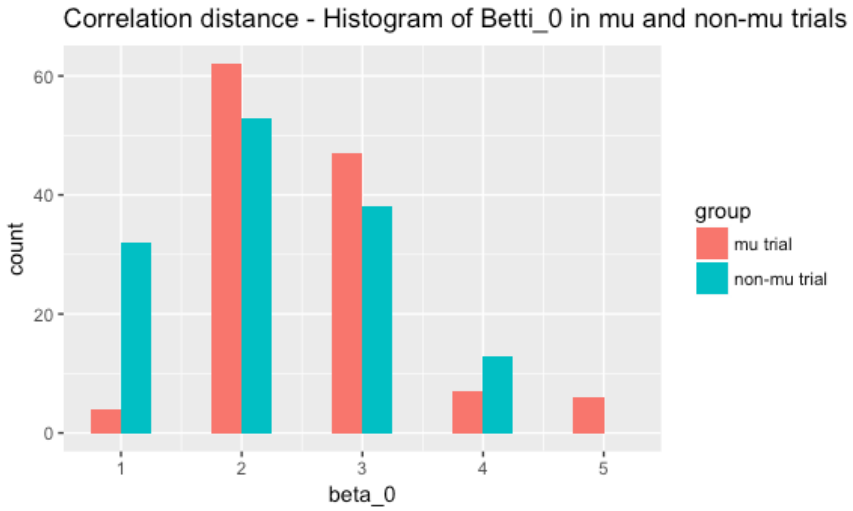

Fig. 5. Histogram of β_0 based on the correlation-metric defined in [7]. The *red bars* show the mu trials, and the zero Betti numbers have a mean β_0 of 2.60 with standard deviation 0.84. The *green bars* show the non-mu trials, and they have a mean β_0 of 2.24 with standard deviation of 0.92. The Wilcoxon rank-sum test has p-value equal to 0.0025, and the Student-t test comparing the two means has p-value equal to 0.001.

Although these differences are statistically significant due to the large number of trials, the differences are subtle. For example, Figure 5 shows that knowing a trial's β_0 would not be sufficient to reliably predict whether it was a mu or non-mu trial. The distance based on correlations reduces size of the data from 27 x 2001 to a 27 by 27 distance matrix. Namely, the distance is summarizing all the information from time series data with rich structures into pairwise correlations, and this is possibly one reason why we observed only low-dimensional topological structure from the resulting Vietoris-Rips complex. This compression of the LFP information appears to be obscuring all the potential topological insight, and this is why we also tried the Euclidean metric.

With the Euclidean metric, both trials in mu and non-mu group show larger β_0, which corresponds to number of connected components in the data cloud repre-

senting a trial. The histogram of these β_0's is shown in Figure 6. The mu group has an average β_0 of 8.40 and standard deviation 2.32. The non-mu group has an average β_0 of 19.71 and standard deviation 5.67. The Wilcoxon rank-sum test has p-value equal to 9.5×10^{-16}, which suggests the two populations have different distribution and that the Euclidean metric can indeed be used as a way to detect difference in topological structures in mu and non-mu trials. The Student-t test has p-value 1.5×10^{-31}, so we can clearly reject the null hypothesis of equal means. Our findings suggest that the data from the mu trials "clusters" more, in the sense that it forms fewer separate connected components.

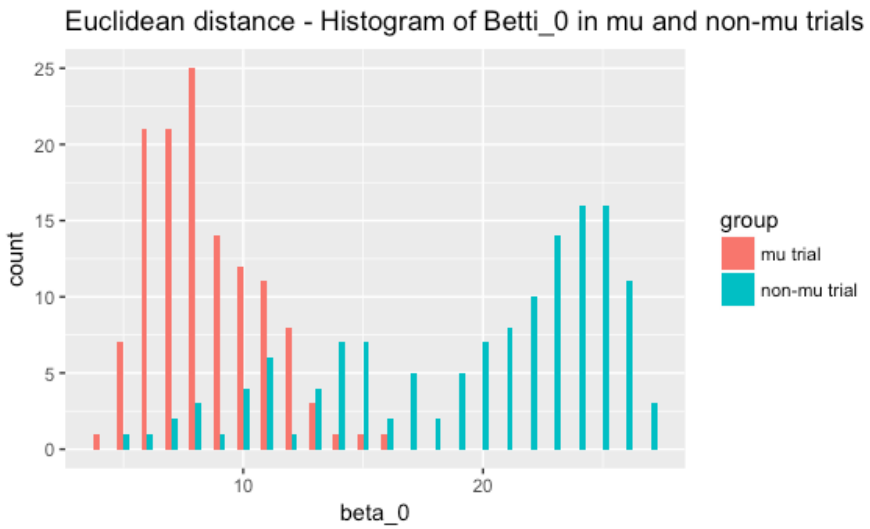

Fig. 6. Histogram of β_0 based on Euclidean distance. The *red bars* show the mu trials, with mean 8.40 and standard deviation 2.33. The *green bars* show the non-mu trials, with mean 19.71 and standard deviation 5.67. The Wilcoxon rank-sum test comparing the two groups has a p-value of 9.5×10^{-16}, which mean we can safely reject the hypothesis that the two populations are from the same distribution. The Student-t test comparing the means returns the p-value of 1.5×10^{-31}.

We also calculated β_1 for each trial, which is number of essential holes in the point-cloud data. Unfortunately this is not as illuminating as the β_0 data in terms of detecting mu trials: 19 out of 126 mu trials have β_1 equal to 1 and, for the non-mu trials, 2 out of 136 have $\beta_1=1$ and one has $\beta_1=2$.

5. Conclusion

In order to test whether a topological analysis can capture differences between distinct brain states as measured by LFPs in awake rats, we compared Betti numbers for segments of multichannel LFP data recorded during an oscillatory "mu" state

and a relatively desynchronized "non-mu" state. A Euclidean-based analysis found Betti-zero numbers in the mu state less than half their values in the non-mu state (Figure 6), reflecting greater clustering of the data cloud in the non-mu state, and supporting that topological analysis can detect functional states of the brain in multichannel LFP data.

In the future, we would like to apply topological data analysis to more sessions and explore other metrics for defining simplicial complexes. It will be interesting to see whether the Betti numbers can capture more subtle functional differences in brain state than those we examined in this study, and whether higher-order Betti numbers can also be useful for distinguishing functional brain states.

Acknowledgments The authors would like to thank the Wellesley College Science Center Summer Research Program and the Brachman-Hoffman Fellowship. Ismar Volić would also like to thank the Simons Foundation for its support. Michael Wiest's work was supported by National Science Foundation Integrative Organismal Systems grants 1121689 and 1353571.

References

[1] Carlsson, G. (2009) Topology and data. *Bulletin of the American Mathematical Society*. 46: 255-308.
[2] Curto, C. (2016) What can topology tell us about the neural code? *Bulletin of the American Mathematical Society*. 54: 63-78.
[3] Edelsbrunner, H.,and Harer, J. (2009) Computational Topology: An Introduction. *American Mathematical Society*.
[4] Fontanini A, Katz DB (2005) 7 to 12 Hz activity in rat gustatory cortex reflects disengagement from a fluid self-administration task. *J Neurophysiol* 93:2832-2840.
[5] Hatcher, A. (2001) Algebraic Topology. *Cambridge University Press*.
[6] Imada A, Morris A, Wiest M (2013) Deviance detection by a P3-like response in rat posterior parietal cortex. *Front Integr Neurosci* 6:127.
[7] Khalid, A., Kim, B.S., Chung, M.K., Ye, J.C., and Jeon, D. (2014) Tracing the evolution of multi-scale functional networks in a mouse model of depression using persistent brain network homology. *NeuroImage*. 101: 351-363.
[8] Munkres, J. (2000) Topology (2nd Edition). *Pearson*.
[9] Nicolelis MA, Baccala LA, Lin RC, Chapin JK (1995) Sensorimotor encoding by synchronous neural ensemble activity at multiple levels of the somatosensory system. *Science* 268:1353-1358.
[10] Nicolelis MA, Fanselow EE (2002) Thalamocortical [correction of Thalamcortical] optimization of tactile processing according to behavioral state. *Nat Neurosci* 5:517-523.
[11] Polack PO, Charpier S (2006) Intracellular activity of cortical and thalamic neurons during high-voltage rhythmic spike discharge in Long-Evans rats in vivo. *J Physiol* 571:461-476.
[12] Rodgers KM, Dudek FE, Barth DS (2015) Progressive, Seizure-Like, Spike-Wave Discharges Are Common in Both Injured and Uninjured Sprague-Dawley Rats: Implications for the Fluid Percussion Injury Model of Post-Traumatic Epilepsy. *J Neurosci* 35:9194-9204.
[13] Shaw FZ (2007) 7-12 Hz high-voltage rhythmic spike discharges in rats evaluated by antiepileptic drugs and flicker stimulation. *J Neurophysiol* 97:238-247.

[14] Tausz, A., Vejdemo-Johansson, M., and Adams, H. (2011) Javaplex: A research software package for persistent (co)homology. *Software available at http://code.google. com/javaplex.*

[15] Vergnes M, Marescaux C, Depaulis A, Micheletti G, Warter JM (1987) Spontaneous spike and wave discharges in thalamus and cortex in a rat model of genetic petit mal-like seizures. *Exp Neurol* 96:127-136.

[16] Wiest MC, Nicolelis MA (2003) Behavioral detection of tactile stimuli during 7-12 Hz cortical oscillations in awake rats. *Nat Neurosci* 6:913-914.