Pascal Grange*

Topology of the mesoscale connectome of the mouse brain

https://doi.org/10.1515/cmb-2020-0106
Received June 30, 2020; accepted September 25, 2020

Abstract: The wiring diagram of the mouse brain has recently been mapped at a mesoscopic scale in the Allen Mouse Brain Connectivity Atlas. Axonal projections from brain regions were traced using green fluoresent proteins. The resulting data were registered to a common three-dimensional reference space. They yielded a matrix of connection strengths between 213 brain regions. Global features such as closed loops formed by connections of similar intensity can be inferred using tools from persistent homology. We map the wiring diagram of the mouse brain to a simplicial complex (filtered by connection strengths). We work out generators of the first homology group. Some regions, including nucleus accumbens, are connected to the entire brain by loops, whereas no region has non-zero connection strength to all brain regions. Thousands of loops go through the isocortex, the striatum and the thalamus. On the other hand, medulla is the only major brain compartment that contains more than 100 loops.

Keywords: Computational topology, connectome, mouse neuroanatomy

MSC: 55U10 (simplicial sets and complexes), 62-07 (data analysis)

1 Introduction

The Allen Mouse Brain Connectivity Atlas [23] has filled a major gap in the knowledge of neuroanatomy by providing a brain-wide map of the connectome of the mouse brain. Adeno-associated viral vectors expressing enhanced green fluorescent protein were injected into the mouse brain, allowing to trace axonal projections. The scale of these experiments is mesoscopic [6], in the sense that injections targeted groups of neurons belonging to brain regions thought to be homogeneous (the microscopic scale would correspond to mapping individual synapses where neurons make contact). The resulting three-dimensional fluorescent traces were registered to brain regions defined in the hierarchical Allen Reference Atlas [11].

These results allowed the construction of the first inter-region connectivity model of the mouse brain. It takes the form of a connectivity matrix, whose rows and columns correspond to a brain region, and whose entries model the connection strengths between pairs of regions defined by classical neuroanatomy. The construction of the model from injection data (possibly infecting several regions for each injection) assumes that projections are homogeneous in each region, and that projections from different sources add to produce projection density to a given region. The connection strengths were observed to span a 10^5-fold range, and are approximately fit by a log-normal distribution.

Some features of the wiring diagram such as node degree and clustering coefficient have been observed to be reproduced by scale-free networks [4] and small-world networks [33], but none of these models was found to fit all observed features of the graph. This begs for further modelling taking into account the global properties of the connectivity map, together with the large range of intensities mapped by connection strengths. On the other hand, the recent computational developments of algebraic topology [13, 12, 34, 32] have given rise to spectacular applications to data analysis in biological sciences, including obstructions to phylogeny in viral evolution [7], and brain networks and neural correlations [10, 14, 28, 27]. Algebraic topology is a branch

***Corresponding Author: Pascal Grange:** Xi'an Jiaotong-Liverpool University, Department of Physics, Suzhou, China, E-mail: Pascal.Grange@xjtlu.edu.cn

of mathematics characterizing properties of spaces that are invariant by continuous deformation, in terms of algebraic quantities.

We will first review the presentation of the Allen Mouse Brain Connectivity Atlas in the form of a matrix modelling connection strengths between brain regions. We will define a mapping from these data to a filtered simplicial complex and explain why the generators of the first homology group are in one-to-one correspondence with closed loops in the mesoscale connectome. The anatomy of these loops will then be analysed by grouping loops by the combinations of major brain compartments that they intersect, and by comparing the fraction of the brain reached from a given region by loops, to the fraction reached by direct connections.

2 Methods

2.1 The matrix of inter-region connectivity strengths in the mouse brain at mesoscopic scale

Each image series in the Allen Mouse Connectivity Atlas corresponds to an injection of tracer, possibly affecting several brain regions. In [23], registered data from 469 image series were combined in order to estimate an inter-region connectivity matrix, presented in matrix form[1]:

$$C(r, r') = \left\{\text{connection strength from region labelled } r \text{ to region labelled } r'\right\}, \quad 1 \le r, r' \le R. \tag{1}$$

The size of this connectivity matrix is $R = 213$, and each of the indices in $\{1, \ldots, R\}$ corresponds to a brain region, defined in the Allen Reference Atlas (ARA, [11, 31]). The derivation of the entries of C assumed homogeneity of brain regions and additivity of connection strengths. Projection densities correspond to axons highlighted by tracers, hence the projection densities from several different sources sum to produce projection density in a given region. This additivity assumption was used to address experimental data in which an injection of viral tracer infected several neighbouring brain regions.

The connectivity matrix captures the local structure of the connectome at mesocopic scale, as it estimates the strength of direct connections from a region to other regions. To capture global features of the connectome, we would like to identify closed circuits constructed from these connections. Algebraic topology formalises this notion: an irreducible closed loop in a topological space (a loop that cannot be shrunk to a point by continuous deformations) is a loop that is not the boundary of a disc in the topological space. It is therefore a one-dimensional cycle that is not the boundary of a two-dimensional object. The family of such objects in a topological space is invariant by continuous deformation, and has a group structure. It is called the first homology group of the topological space, and denoted by H_1.

If we repeat this reasoning in dimension zero, we obtain the more familiar notion of connected component: elements of H_0 are zero-dimensional objects that cannot be joined by a path drawn on the topological space. They correspond to distinct connected components. More generally, the elements of the homology group H_k are objects of dimension k that are not boundaries of $(k + 1)$-dimensional objects in a topological space, and therefore formalise the notion of hole. Moreover, the algebraic structure of homology groups allows us to count independent objects in each of them. We can write H_k as the sum of b_k copies of $\mathbf{Z}$ (and possibly finite groups). The number b_k is called the k-th Betti number of the topological space: it is the number of independent generators of infinite subgroups of the k-th homology group. The Betti numbers b_0 and b_1 are the number of connected components and the number of independent loops respectively.

1 The entries of the matrix C correspond to the left-hand side of Fig. 3 in [23].

2.2 Mapping the matrix of connection strengths to a filtered simplicial complex

The connectivity matrix C is not symmetric, because projections from a given brain region to other brain regions correspond to axons and are therefore oriented. If we map the connectivity matrix to a weighted graph with R vertices corresponding to brain regions, and weighted links corresponding to the entries of the connectivity matrix, links with two different orientations and different weights can exist between pairs of vertices. Let us apply a decreasing function to the entries of the connectivity matrix and define for instance:

$$d(r, s) = -\log\left(C(r, s) + \epsilon\right), \tag{2}$$

where ϵ is a positive regulator chosen to be smaller than the minimum connection strength, so that entries of d are bounded (and r and s are two integers in $[1..R]$ labelling brain regions). Persistent loops are independent of the choice of ϵ (in practice we took $\epsilon = 0.01\min_{r,s} C(r, s)$, resulting in a maximum entry of 42 for d). This operation results in an approximately normal distribution of entries in the matrix d. The quantity $d(r, s)$ has been observed in [23] to be positively correlated to the spatial distance between region labelled r and region labelled s. However, the entries of d are not quite distances since d is not symmetric.

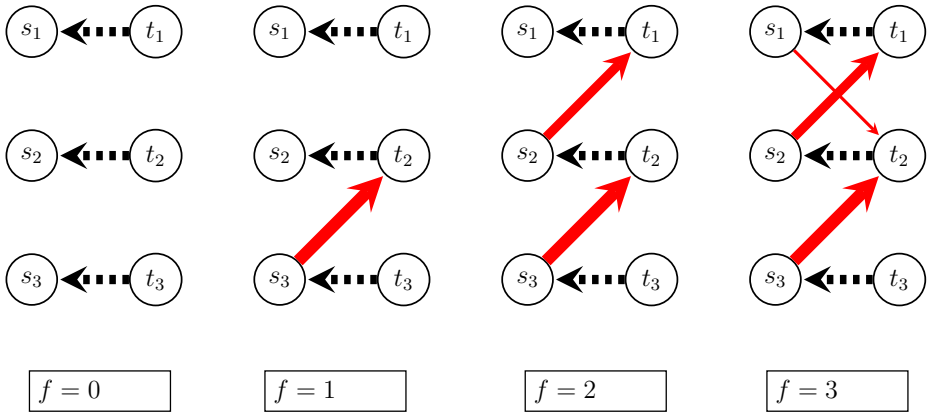
We propose to map brain regions to a graph with $2R$ vertices as follows. Each brain region, labelled r, is mapped to a pair of vertices (s_r, t_r), where s_r represents the sources of action potentials (the axons of the cell bodies in the region, that conduact axon potentials), and t_r the targets of action potentials (the dendrites in the region, that receive action potentials). A link is drawn between vertices t_r and s_r for each r in $\{1, \ldots, R\}$. With this doubling prescription we declare links between targets and sources to be present in each region, and we are going to look for closed paths constructed from the non-zero entries $C(s_r, t_r)$ of the connection matrix. Direct approaches to persistent homology for asymmetric networks have been proposed, based on the Dowker complex, see [8], but the present doubling prescription takes advantage of the mesoscopic scale of the connectome data.

Given the family of $2R$ labelled vertices we have just described, we can construct a family of graphs and work out the families of independent loops for each of them by executing the following pseudo-code:

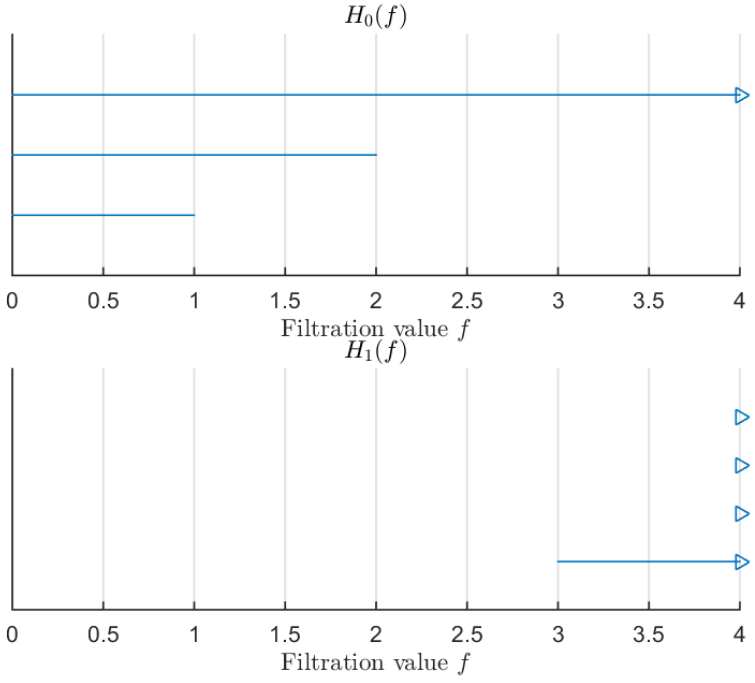
1. Consider a fixed $f \in]0, \max_{1\leq r<s\leq R} d(r, s)]$ (referred to as a filtration value).
2. Draw a line between any source labelled s_a and any target labelled t_b such that $d(s_a, t_b) \leq f$.
3. Work out generators of the homology groups $H_0(f)$ (connected components) and $H_1(f)$ (independent loops) in the resulting graph.

The third step uses techniques of simplicial homology, implemented in JavaPlex [2, 1]. Simplices in our case consist of vertices (zero-simplices, corresponding to brain regions), and links (one-simplices, corresponding to axons connecting two brain regions). The result of the procedure depends on the value of the filtration value f. At $f = 0$ we have as many connected components as points, and no loop. The number of connected components decreases when f increases. It reaches 1 for some value of the filtration value. When f is increased beyond this value, new connections can be drawn, possibly changing the number of loops. The growing family of graphs depending on the parameter f is called a simplicial complex. The results of the above procedure are sketched in Fig. 1 for a toy model consisting of three regions (and connection strengths assumed to be mapped to $d = \begin{pmatrix} 0 & 3 & 4 \\ 2 & 0 & 4 \\ 4 & 1 & 0 \end{pmatrix}$). The pseudo-code is executed for each unique value of the entries of d, in ascending order.

When f reaches the maximum entry of the matrix d, the simplicial complex cannot be changed anymore by increasing the filtration value. If we repeat the above procedure for all values of the entries of the matrix d, arranged in increasing order, we can work out for each feature F (which can be a connected component or a loop) an interval of filtration values $[f_{min}(F), f_{max}(F)]$ for which the feature F is present. These intervals can be drawn for all existing features F of a fixed dimension, yielding graphs of barcode form, one per dimension of feature (see Fig. 1B). Features persistent over a longer interval are thought less likely to be due to noise [9].



(a)



(b)

Figure 1: A toy model with three regions. (A) Each of the regions is doubled at filtration value zero, resulting in a source vertex and a target vertex. The filtered complex start with three connected components and no loop. For increasing using values of the filtration parameters, links are drawn from to reflect projections from the nodes in the family $(s_r)_{1\leq r\leq 3}$ to nodes in the family $(s_r)_{1\leq r\leq 3}$. (B) The graph gets connected at filtration value $f = 2$ (resulting in $b_0(f) = 1$ for $f \geq 2$). A loop appears at $f = 3$, the corresponding generator of the first homology group is $[s_1, t_2] + [t_2, s_2] + [s_2, t_1] + [t_1, s_1]$. It goes through the regions labelled 1 and 2.

In particular, we will restrict ourselves to cycles that persist from their first appearance (at a given filtration value) to the maximum of d.

2.3 Neuroanatomy of cycles

2.3.1 Three-dimensional presentation of brain regions

To analyse the anatomy of persistent cycles, we used the voxelised version of the Allen Reference Atlas at a spatial resolution of 25 microns [2]. This three-dimensional grid contains numerical labels encoding neuroanatomy at the finest level compatible with the resolution (the voxelised atlas consists of a grid of $V_{tot} \simeq 3.1 \times 10^7$ cubic voxels, with 677 distinct regions). These numerical labels are associated to the names of brain regions accroding to the hierarchical annotation of the ARA [3]. For each of the $R = 213$ regions in the mesoscale model of the connectome (Eq. 1), we resolved the numerical label and the corresponding voxels. If the region has descendants in the hierarchical annotation, we lump together the voxels corresponding to these descendants, by annotating them with the numerical label of the region. After this step, the voxelised atlas contains 361 distinct regions. The regions included in the mesoscale model of the connectome therefore do not quite span the entire brain. Let us denote by V_r the number of voxels in region labelled r:

$$V_r := |\{v, \text{ voxel labelled } v \in \text{ region labelled } r\}|, \quad 1 \leq r \leq R. \tag{3}$$

The brain regions corresponding to the rows of the connectivity matrix contain

$$V := \sum_{r=1}^{R} V_r \simeq 2.4 \times 10^7 \text{ voxels}, \tag{4}$$

about 79 percent of the total volume V_{tot} of the brain. In the rest of this paper we will disregard the brain regions that are not included in the matrix of connectivity strengths, and the volume V will be referred to as the volume of the brain.

Brain regions are arranged hierarchically [11, 31]. Grouping brain regions at a coarse hierarchical level yields the following major brain compartments which will sometimes be designated by acronyms for presentation purposes: Isocortex, olfactory areas (OLF), cortical subplate (CTXsp), striatum (STR), pallidum (PAL), thalamus (TH), hypothalamus (HY), midbrain (MB), medulla (MY), cerebellar cortex (CBX). These major brain compartments will be referred to as *big regions*.

2.3.2 Fraction of the brain connected by loops to a given region

Given a brain region labelled r and a filtration value f, let us denote by $\mathcal{C}_r(f)$ the family of generators of the first homology group at filtration value f that go through the region r:

$$\mathcal{C}_r(f) := \left\{c \in H_1(f), r \in c\right\}, \tag{5}$$

and by $\overline{\mathcal{C}_r(f)}$ the set of vertices through which the elements of $\mathcal{C}_r(f)$ go:

$$\overline{\mathcal{C}_r(f)} = \left\{s \in \{1, \ldots, R\}, \exists c \in H_1(f), s \in c\right\}. \tag{6}$$

In the toy model of Fig. 1 with three brain regions, the sets $\mathcal{C}_1(2)$ and $\mathcal{C}_3(3)$ are empty. The set $\mathcal{C}_1(3)$ consists of one loop, corresponding to $\overline{\mathcal{C}_r(f)} = \{1, 2\}$.

Given the voxelised annotation of the brain, we can calculate the volume of the regions connected to the region r by loops, as a fraction of the total number of voxels covered by the regions included in the mesoscale connectome:

$$\phi_r(f) = \frac{1}{V} \sum_{s \in \overline{\mathcal{C}_r(f)}} V_s, \tag{7}$$

---

2 See download instructions and code snippets in MATLAB and Python on the Allen Brain Atlas data portal:
`http://help.brain-map.org/display/mouseconnectivity/API#API-DownloadAtlas3-DReferenceModels`

3 The hierarchical system of annotation containing names of brain regions and numerical ids is available online at `http://api.brain-map.org/api/v2/structure_graph_download/1.json`

where V is the total number of voxels defined in Eq. 4, and V_s is the number of voxels in region labelled s. As the volumes of brain regions are not all equal, we can define an aletrantive measure by the fraction of the total number of regions connected to region labelled r by loops:

$$\nu_r(f) = \frac{1}{R}\left|\overline{\mathcal{C}_r(f)}\right|. \tag{8}$$

For a region labelled r that is connected by loops to all brain regions, for some value of the filtration f, we will have $\phi_r(f) = \nu_r(f) = 1$.

It is natural to compare these ratios to the analogous ratios obtained by taking into account only direct connections between pairs of brain regions, as encoded by the connectivity matrix C. Let us denote these quantities by $\phi_r^{(c)}(f)$ for the fraction of volume and by $\nu_r^{(c)}(f)$ for the fraction of the number of regions:

$$\phi_r^{(c)}(f) = \frac{1}{R}\sum_{s=1}^{R} V_s \max\left(\mathbf{1}(C(r,s) > f), \mathbf{1}(C(s,r) > f)\right), \tag{9}$$

$$\nu_r^{(c)}(f) = \frac{1}{R}\sum_{s=1}^{R} \max\left(\mathbf{1}(C(r,s) > f), \mathbf{1}(C(s,r) > f)\right), \tag{10}$$

where $\mathbf{1}$ denotes the indicator function. By construction there can be regions in the family $\overline{\mathcal{C}_r(f)}$ that are not directly connected to region r based on the inter-region connectivity matrix. There can also be regions with projections to or from region r that are not involved in any closed cycle going through region r at the given filtration value f. There is therefore no a priori solidarity between the quantities ϕ and ν based on cycles and their analogues $\phi^{(c)}$ and $\nu^{(c)}$ based on direct connections.

2.4 Brain-wide maps of loops going through a given brain region

To sum up the connections by loops between two given brain regions, we can calculate the sum of connection strengths at which cycles connecting these two regions appear, weighted by the numbers of such cycles. As the generators of the first homology groups are independent, this is consistent with the additivity assumption of connection strengths between brain regions. For two regions labelled r and s we introduce

$$\omega(r,s) := \sum_f e^{-f}\mathbf{1}(s \in \overline{\mathcal{C}_r(f)}), \tag{11}$$

where the sum is over the filtration values, and the exponential function comes from Eq. 2 relating connection strengths to the distance used to defined filtration values.

Fixing the region label r and allowing the region label s to take all the possible values in $\{1,\ldots,R\}$, we can define a brain-wide map connection strength based on cycles to region labelled r, by defining for voxel labelled v:

$$\mathcal{W}_r(v) = \sum_f e^{-f}\mathbf{1}(s(v) \in \overline{\mathcal{C}_r(f)}), \tag{12}$$

where $s(v)$ is the index (in {1,…,R}) of the brain region to which voxel labelled v belongs. The quantity $\mathcal{W}_r$ maps a voxel to a real number, and can be visualised as a heat map for instance.

To estimate how localised a cycle is in space, we can compute the Kullback–Leibler divergence from the above-defined profile and the uniform brain-wide density. If we denote by the label of the region to which voxel labelled v belongs, this divergence is expressed as

$$KL(r) = \frac{1}{V}\sum_{v=1}^{V} W_r(v)\log(W_r(v)), \tag{13}$$

where V is the total number of voxels defined in Eq. 4, and we used the fact that W_r/V is a normalised density function. This divergence induces a ranking of brain regions.

3 Results

3.1 Persistent cycles

The bar code corresponding to the persistent generators of the first two homology groups for the filtered ipsilateral complex is shown on Fig. 2. The maximum Betti numbers are too large for the individual bars to be visible, as they are on Fig. 1B. By construction the Betti number b_0 starts at $R = 213$, and the plateau at filtration value 25 in the bar code of the first homology group corresponds to $b_1 = 16,529$. At filtration values larger than $f_{conn} = 7$, all brain regions are in the same connected component based. Moreover, from the growth profile of the bar code of the first homology group, we can see that thousands of persistent cycles appear at filtration values lower than f_{conn} ($b_1 = 8,230$ at $f = f_{conn}$).

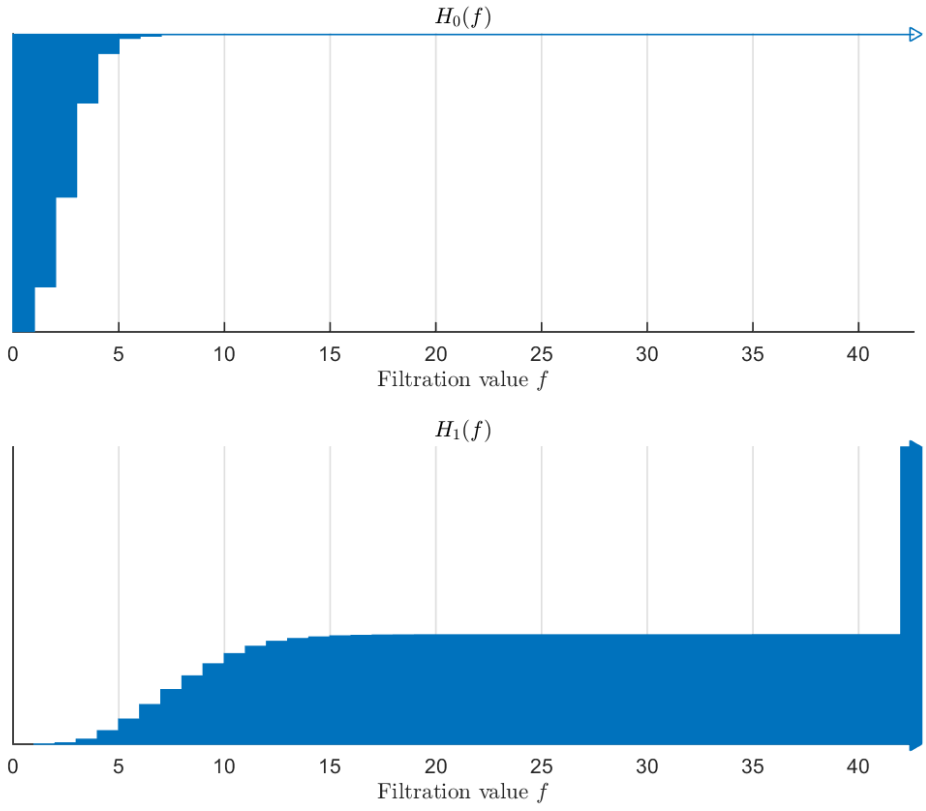


Figure 2: Bar code of the complex based on the inter-region connectivity matrix. The second plateau that appears at filtration value $f = 42$ corresponds to $-\log \epsilon$, where ϵ is the regulator introduced in Eq. 2. The shape of the barcode and the families of generators of H_1 and H_0 would be unchanged if the regulator went to zero.

3.2 Loops highlight the cortico-striato-thalamic network

To obtain a coarse picture of the anatomy of cycles, let us work out which family of big regions (as defined in Section 2.3.1) is intersected by each loop, at a given filtration value. We can rank the resulting families of big regions by decreasing value of prevalence. For an example of the brain regions in a loop appearing at filtration value $f = 2$, together with the names of the corresponding big regions, see Table 1.

Table 1: Brain regions in a circuit appearing at filtration value $f = 2$. Six regions appear, belonging to four distinct *big* brain regions.

Brain region	Big region
Anterior cingulate area, dorsal part	Isocortex
Primary auditory area	Isocortex
Caudoputamen	Striatum
Medial geniculate complex, dorsal part	Thalamus
Medial geniculate complex, ventral part	Thalamus
Superior colliculus, motor related	Midbrain

For definiteness we can do this calculation at two special filtration values:
- the filtration value $f_{conn} = 7$ at which the Betti number b_0 becomes 1.
- the maximum filtration value ($f_{max} = 37$), which implies that all loops are taken into account.
We found that 334 distinct families of big regions occur, but the top 27 families account for 50% of the loops. The most prevalent families of big regions are presented in Table 2, where they are ordered by decreasing prevalence at filtration value $f = 7$. We notice that midbrain is represented in most of the rows of the table[4], and that the most frequent combinations go through six or more big regions. The family of four big regions corresponding to the regions in the loop presented in Table 1 is the fourth most prevalent family of big regions in loops appear at filtration values lower than $f = 7$, with 249 distincth loops (resp. fifth, $f = 37$ and 328 loops). However, if we focus on small families of big regions (with three regions or fewer), the most prevalent is the cortico-striato-thalamic family (7th row of the table, with 183 loops at $f = 7$). Highlighting occurrences of isocortex, striatum and thalamus in colour in Table 1 (red, green and blue respectively), we notice that the first 17 rows present at least one of these colours (often associated with midbrain, highlighted in brown). The next small family consists of isocortex and striatum (16th rank, with 106 loops), followed at rank 18 by medulla. With 104 loops, medulla contains more cycles than the cortico-thalamic family (ranked 25 with 80 loops). Moreover, medulla is the only big region to support more than 10 cycles (see Table 3, which shows that only 147 cycles are confined to a single big region, out of which 134 appear before the filtration value 7).

To assess whether the dominance of medullar loops among loops confined to a single big region could have been guessed based on the matrix of connectivity strengths, let us calculate how much of the connection strengths from its sub-regions project within the same region. For each big region labelled b, let us define:

$$\frac{\mathcal{C}_{intra}}{\mathcal{C}_{tot}}(b) = \frac{\sum_{r=1}^{R}\sum_{s=1}^{R}\mathbf{1}(B(r)=b)\mathbf{1}(B(s)=b)C(r,s)}{\sum_{r=1}^{R}\sum_{s=1}^{R}\mathbf{1}(B(r)=b)C(r,s)}. \tag{14}$$

The ratio $\mathcal{C}_{intra}/\mathcal{C}_{tot}$ is the conditional probability of a connection from within the big region b to project within region b. The values are plotted on Fig. 3. The average fraction is 31%, and the values range from 6% (thalamus) to 68% (cerebellar cortex, which is found to contain only one loop, see Table 3B).
The fact that 80 percent of the cycles confined to a single big region are in medulla could not have been guessed based on the connectivity matrix only: 47% of connections from medullar regions project to medullar regions, which is above average, but the medulla is only ranked fourth among 12 big regions by our measure of conditional probability.

3.3 Persistent loops can connect brain regions to the entire brain

On average, loops connect a given brain region to a larger domain in the brain than direct projections. This is true for both the volumetric and the counting measures defined in Eqs 9 and 10, and at all filtration values.

4 Midbrain is also represented in most of the loops that appear at low filtration values $f < 3$.

Table 2: Table of combinations of big regions supporting loops. The combinations are ordered by decreasining number of cycles supported (at filtration value $f = 7$).

Rank (out of 334)	Number of loops ($f = 7$)	Cumulated fraction of loops ($f = 7$)	Number of loops ($f = 37$)	Cumulated fraction of loops ($f = 37$)	Big regions intersected
1	410	6%	600	4%	Cortical Subplate, Hippocampal Formation, Hypothalamus, Isocortex, Midbrain, Striatum, Thalamus
2	268	9%	614	8%	Cortical Subplate, Hypothalamus, Isocortex, Midbrain, Striatum, Thalamus
3	265	12%	455	11%	Cortical Subplate, Hypothalamus, Isocortex, Midbrain, Pallidum, Striatum, Thalamus
4	249	15%	328	13%	Isocortex, Midbrain, Striatum, Thalamus
5	248	19%	346	15%	Hypothalamus, Isocortex, Midbrain, Pallidum, Striatum, Thalamus
6	199	21%	283	17%	Hypothalamus, Isocortex, Midbrain, Striatum, Thalamus
7	183	23%	222	19%	Isocortex, Striatum, Thalamus
8	165	25%	255	20%	Cortical Subplate, Isocortex, Midbrain, Striatum, Thalamus
9	141	27%	175	21%	Hypothalamus, Isocortex, Midbrain, Thalamus
10	135	29%	175	22%	Hypothalamus, Midbrain, Pallidum, Thalamus
11	130	31%	190	24%	Cortical Subplate, Hippocampal Formation, Hypothalamus, Midbrain, Striatum, Thalamus
12	130	32%	208	25%	Cortical Subplate, Hippocampal Formation, Hypothalamus, Isocortex, Midbrain, Pallidum, Striatum, Thalamus
13	122	34%	149	26%	Cortical Subplate, Isocortex, Striatum
14	115	35%	163	27%	Hypothalamus, Isocortex, Midbrain, Pallidum, Thalamus
15	107	37%	135	28%	Cortical Subplate, Hypothalamus, Isocortex, Midbrain, Pallidum, Striatum
16	106	38%	117	29%	Isocortex, Striatum
17	105	39%	139	29%	Cortical Subplate, Hippocampal Formation, Hypothalamus, Midbrain, Olfactory Areas, Striatum, Thalamus
18	104	40%	115	30%	**Medulla**
19	96	42%	108	31%	Hypothalamus, Midbrain, Thalamus
20	85	43%	104	32%	Hypothalamus, Midbrain, Pallidum
21	85	44%	111	32%	Hypothalamus, Isocortex, Midbrain, Pallidum, Striatum
22	84	45%	118	33%	**Medulla**, Midbrain, Pallidum

This can be observed on Fig. 4, where the following averages across all brain regions are plotted:

$$\langle \phi(f) \rangle := \frac{1}{V} \sum_{i=1}^{R} \phi_r(f), \quad \langle \nu(f) \rangle := \frac{1}{R} \sum_{i=1}^{R} \nu_r(f), \quad \langle \phi^{(c)}(f) \rangle := \frac{1}{R} \sum_{i=1}^{R} \phi_r^{(c)}(f), \quad \langle \nu^{(c)}(f) \rangle := \frac{1}{R} \sum_{i=1}^{R} \nu_r^{(c)}(f). \tag{15}$$

Moreover, all four averages reach an asymptote, and the ratios of asymptotic values do not depend heavily on the choice of measure, as:

$$\langle \phi(25) \rangle \simeq 1.39 \langle \phi^{(c)}(25) \rangle \simeq 87\%, \quad \text{and} \quad \langle \nu(25) \rangle \simeq 1.37 \langle \nu^{(c)}(25) \rangle \simeq 80\%. \tag{16}$$

Table 3: Tables of single big regions containing loops, ordered by decreasing number of loops. (A) For filtration value $f = 7$, there is a total of 134 cycles. (B) For filtration value $f = 37$, there are 147 loops, and just one more region represented.

Number of loops	Fraction (%)	Big region
104	78	Medulla
8	6	Hippocampal Formation
8	6	Hypothalamus
6	5	Olfactory Areas
5	4	Isocortex
3	3	Midbrain

(A)

Number of loops	Fraction (%)	Big region
115	79	Medulla
9	7	Hypothalamus
8	6	Hippocampal Formation
6	5	Olfactory Areas
5	4	Isocortex
3	3	Midbrain
1	1	Cerebellar Cortex

(B)

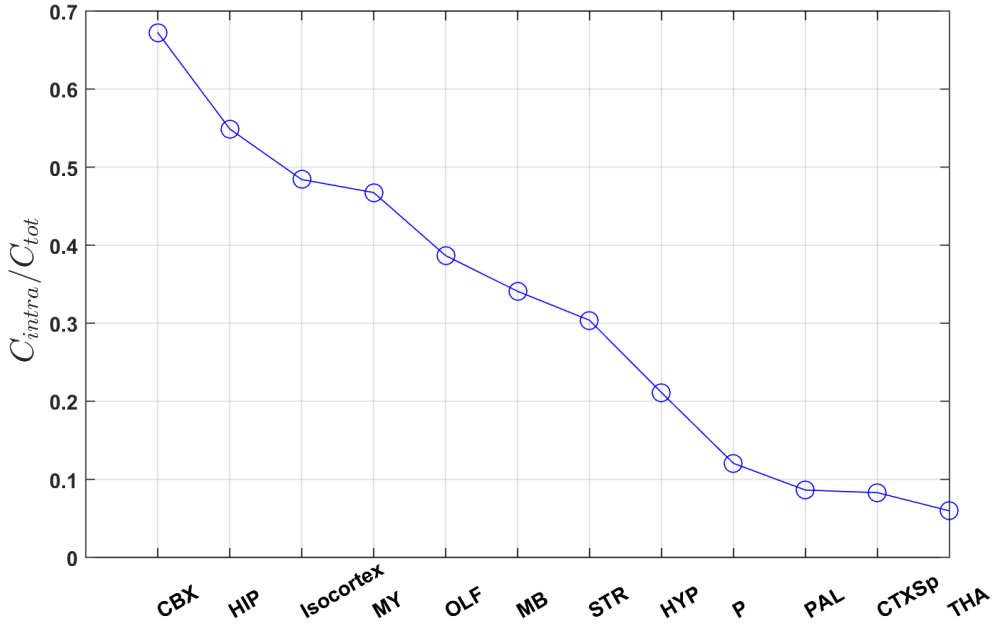


Figure 3: Fraction of inter-region connection strengths supported in the big region of origin. Acronyms of big regions read as follows: OLF = olfactory areas, HPF = hippocampal formation, CTXsp = cortical subplate, STR = striatum, PAL = pallidum, TH = thalamus, HY = hypothalamus, MB = midbrain, MY = medulla, CBX = cerebellar cortex.

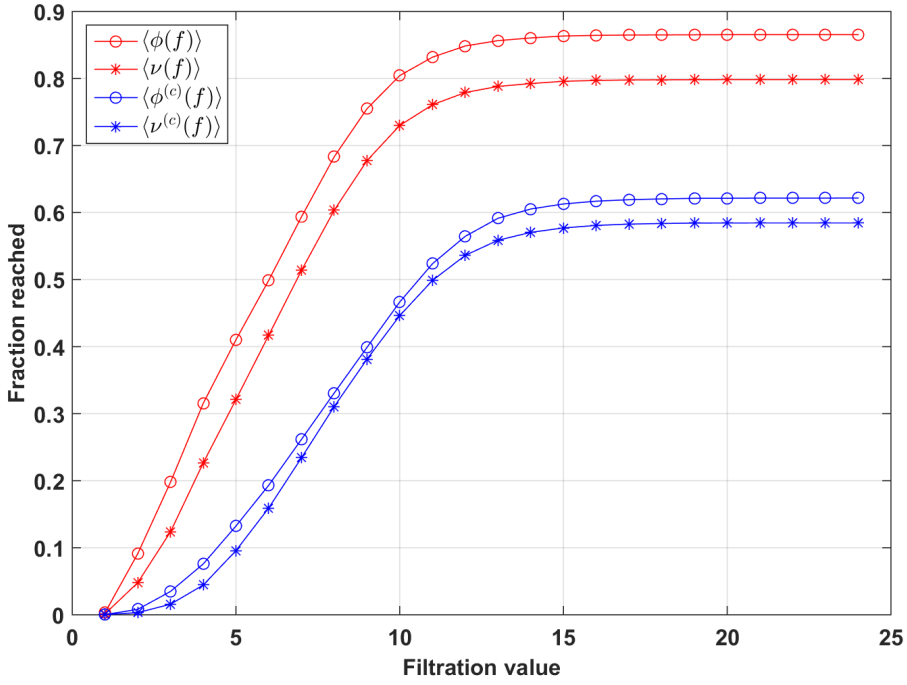


Figure 4: Fraction of the brain reached by loops, averaged across all regions in the connectivity atlas.

However, the maximum values of the fractions ϕ_r and v_r are highly heterogeneous across brain regions. The heat map of Fig. 5 shows the fraction of the brain $\phi_r(f)$, with the region labels r grouped by big regions (and ordered within each big region by decreasing value of the maximum reached at large filtration value). In particular, 18 regions reach the entire brain through connections by loops. Their names are shown on Table 4, together with the big regions to which they belong (midbrain is the most represented big region, with 4 regions).

3.4 Brain-wide density of connections by loops to a given region

The region for which W_r has lowest Kullback–Leibler divergence from a uniform profile is the nucleus accumbens. The region with highest Kullback–Leibler divergence from a uniform profile is the flocculus, a region in the cerebellar cortex. Heat maps of the densities of loops for both regions are shown on Fig. 6. For each of these regions the volume can be presented either as the sum of values of voxels (subfigures (a) and (c)) or as maximal-intensity projections (subfigures (b) and (d)). On maximal-intensity projections of the density W_r, the voxels belonging to region labelled r have the largest value by construction (because all loops go through this region), and maximal-intensity projections can exhibit at most R different values (they are piecewise constant on each brain region). On the other hand, projections of sums of the density W_r can have higher values in voxels that do not belong to region labelled r due to thickness effects.

4 Discussion

The data encoded in the connectivity strengths are local in nature: they encode the regions to which a given region is connected by axonal projections. Closed loops defined by axonal projections in the brain need global

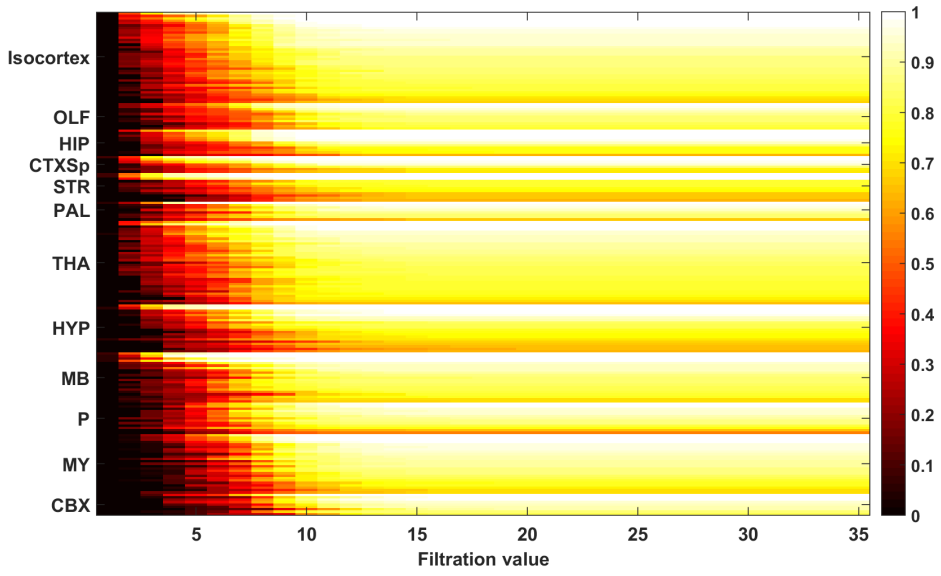
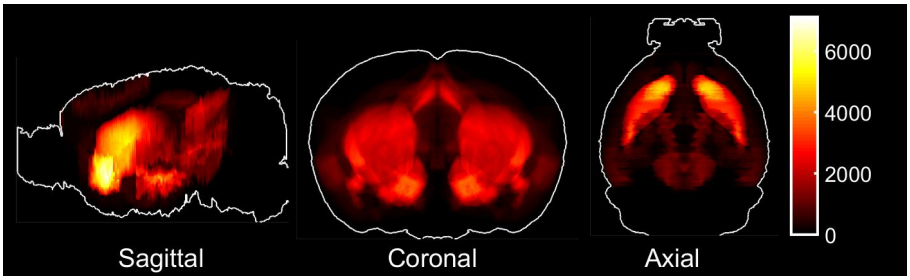


Figure 5: Heat maps of the fraction of the brain reached by loops ϕ_r, as a function of the filtration value f (horizontal axis). The rows correspond to brain regions (grouped by big regions according to coarse neuroanatomy). The subregions in each big region are ordered by decreasing value of the asymptote $\phi_r(35)$.

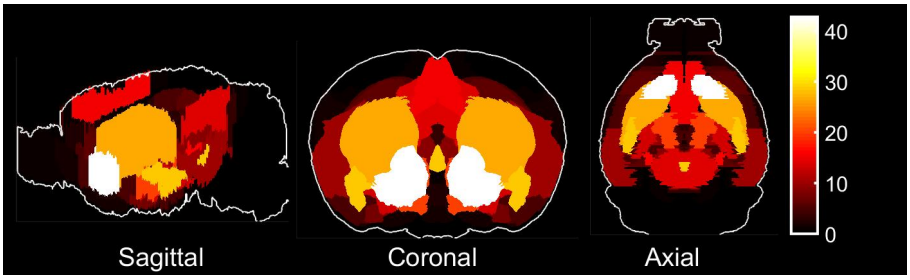
Table 4: Table of brain regions connected to all brain regions by loops. On average these brain regions are connected to 72% of the volume of the brain by direct connections, which is largest than the average value $\langle \phi^{(c)} \rangle \simeq 63\%$. However, 5 regions correspond to values below this average.

Big region	**Name of region**	$\phi_r^{(c)}(25)(\%)$	**Rank by $\phi^{(c)}$, out of** 213
Isocortex	Anterior cingulate area, dorsal part	85	13
Hippocampal Formation	Entorhinal area, lateral part	81	31
Cortical Subplate	Basolateral amygdalar nucleus	81	29
Cortical Subplate	Posterior amygdalar nucleus	50	173
Striatum	Nucleus accumbens	68	73
Striatum	Caudoputamen	55	148
Striatum	Medial amygdalar nucleus	59	123
Pallidum	Globus pallidus, internal segment	77	40
Thalamus	Medial geniculate complex, dorsal part	71	64
Thalamus	Peripeduncular nucleus	74	52
Hypothalamus	Lateral hypothalamic area	70	67
Hypothalamus	Subthalamic nucleus	73	56
Midbrain	Central linear nucleus raphe	75	49
Midbrain	Midbrain reticular nucleus	56	138
Midbrain	Periaqueductal gray	66	87
Midbrain	Superior colliculus, motor related	59	125
Cerebellar Cortex	Ansiform lobule	98	1

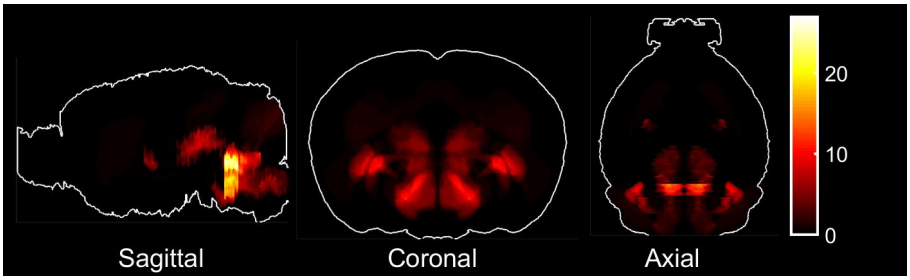
tools to be detected: indeed such loops may connect pairs of regions that are not directly connected by axonal projections (an extreme example of such a situation would be a single loop going through all the regions, in which each region would only project to the next region). Persistent homology addresses this global feature of the problem, moreover it takes into account the many orders of magnitude spanned by the connection strengths by treating connection strengths as filtration values.



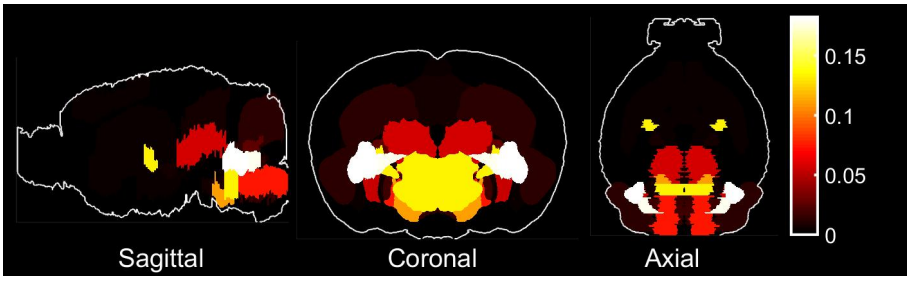
(a)



(b)



(c)



(d)

Figure 6: Heat map of the weighted regions connected by loops to given brain regions, projected in the sagittal, coronal and axial directions. (A) Nucleus accumbens, projection of sums of intensities. (B) Nucleus accumbens, projection of maximal intensities. (C) Flocculus, projection of sums of intensities. (D) Flocculus, projection of maximal intensities. For visualisation purposes, both hemispheres are shown, even though the results are based on the ipsilateral connectivity matrix only.

Persistent homology reveals global properties of the mouse connectome at the mesoscopic scale. The high prevalence of cortico-striato-thalamic loops is a strong consistency check between the mesoscale approach to connectomics [6, 23] and the present topological method, as the cortico-striato-thalamo-cortical circuit is known to regulate complex behaviours (for a review of its involvement in the Tourette sydrome see [25]). The relative abundance of closed loops in the medulla is perhaps best interpreted in terms of the automatic

functions performed by the medulla, which are topologically insulated from circuits involved in conscious behaviours or learning [26].

Axons, in addition to conducting action potentials, transport biological molecules [24, 19]. Integration with other brain-wide data sets, in particular gene-expression data [20, 22], has already shown that connected regions in the adult rodent brain have increased similarity in gene-expression profiles, and sets of genes most correlated to connectivity have been identified (see [**?**]). It would be interesting to see how these correlations (identified based on connectivity data in the rat brain) persist across the loops that we identified in the mouse brain. Moreover, spatial gene expression patterns have been related to cortico-striatal functional networks [3]. Loops that are as persistent as the cortico-striatal ones yield natural candidates for places where to look for such genetic signatures of functional networks. Moreover, the brain-wide coverage of the gene-expression atlas [20, 22] allows to estimate spatial densities of cell types known by their transcriptional activity [15, 16, 5], which yields neuroanatomical predictions relevant to brain disorders [21, 17]. The present work could lead to insights into the correlation structure between the local density of loops and the spatial density of cell types. Another direction of research is the conservation of patterns across species [29, 30, 18], even though gene-expression and connectivity data do not have the same brain coverage as in the mouse atlas.

We have worked out the persistent loops of the wiring diagram of the brain based on ipsilateral projections only, leaving out the issue of contralateral projections. Contralateral projections have been mapped (see [23], right-hand side of Fig. 4A), but they do not allow to work out closed loops because the missing projections would involve injecting tracers into the left hemisphere, and mapping the contralateral projections again. However, the known contralateral connection strengths have been found to be significantly weaker than ipsilateral ones. One can therefore conjecture than loops going through brain regions in different hemispheres do not dramatically change the global features of the wiring diagram of the brain.

Acknowledgments: This work is supported by the Research Center for Precision Medicine, HT-URC, Xi'an Jiaotong-Liverpool University, Suzhou, China.

Author's statement: Conflict of interest: the author states no conflict of interest.

References

[1] Adams, H. and Tausz, A. (2011). Javaplex tutorial. *Google Scholar*.

[2] Adams, H., Tausz, A., and Vejdemo-Johansson, M. (2014). Javaplex: A research software package for persistent (co) homology. In *International Congress on Mathematical Software*, pages 129–136. Springer.

[3] Anderson, K. M., Krienen, F. M., Choi, E. Y., Reinen, J. M., Yeo, B. T., and Holmes, A. J. (2018). Gene expression links functional networks across cortex and striatum. *Nature communications*, 9(1):1–14.

[4] Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *science*, 286(5439):509–512.

[5] Bohland, J. W., Bokil, H., Pathak, S. D., Lee, C.-K., Ng, L., Lau, C., Kuan, C., Hawrylycz, M., and Mitra, P. P. (2010). Clustering of spatial gene expression patterns in the mouse brain and comparison with classical neuroanatomy. *Methods*, 50(2):105–112.

[6] Bohland, J. W., Wu, C., Barbas, H., Bokil, H., Bota, M., Breiter, H. C., Cline, H. T., Doyle, J. C., Freed, P. J., Greenspan, R. J., et al. (2009). A proposal for a coordinated effort for the determination of brainwide neuroanatomical connectivity in model organisms at a mesoscopic scale. *PLoS Comput Biol*, 5(3):e1000334.

[7] Chan, J. M., Carlsson, G., and Rabadan, R. (2013). Topology of viral evolution. *Proceedings of the National Academy of Sciences*, 110(46):18566–18571.

[8] Chowdhury, S. and Mémoli, F. (2018). A functorial dowker theorem and persistent homology of asymmetric networks. *Journal of Applied and Computational Topology*, 2(1-2):115–175.

[9] Collins, A., Zomorodian, A., Carlsson, G., and Guibas, L. J. (2004). A barcode shape descriptor for curve point cloud data. *Computers & Graphics*, 28(6):881–894.

[10] Curto, C., Giusti, C., Marku, K., Pastalkova, E., and Itskov, V. (2013). Pairwise correlation graphs from hippocampal population activity have highly non-random, low-dimensional clique topology. *BMC neuroscience*, 14(1):1–2.

[11] Dong, H. W. (2008). *The Allen reference atlas: A digital color brain atlas of the C57Bl/6J male mouse.* John Wiley & Sons Inc.

[12] Edelsbrunner, H. and Harer, J. (2010). *Computational topology: an introduction*. American Mathematical Soc.

[13] Edelsbrunner, H., Letscher, D., and Zomorodian, A. (2000). Topological persistence and simplification. In *Proceedings 41st annual symposium on foundations of computer science*, pages 454–463. IEEE.

[14] Giusti, C., Pastalkova, E., Curto, C., and Itskov, V. (2015). Clique topology reveals intrinsic geometric structure in neural correlations. *Proceedings of the National Academy of Sciences*, 112(44):13455–13460.
[15] Grange, P., Bohland, J. W., Okaty, B. W., Sugino, K., Bokil, H., Nelson, S. B., Ng, L., Hawrylycz, M., and Mitra, P. P. (2014). Cell-type–based model explaining coexpression patterns of genes in the brain. *Proceedings of the National Academy of Sciences*, 111(14):5397–5402.
[16] Grange, P., Hawrylycz, M., and Mitra, P. P. (2013). Computational neuroanatomy and co-expression of genes in the adult mouse brain, analysis tools for the allen brain atlas. *Quantitative Biology*, 1(1):91–100.
[17] Grange, P., Menashe, I., and Hawrylycz, M. (2015). Cell-type-specific neuroanatomy of cliques of autism-related genes in the mouse brain. *Frontiers in computational neuroscience*, 9:55.
[18] Hawrylycz, M., Miller, J. A., Menon, V., Feng, D., Dolbeare, T., Guillozet-Bongaarts, A. L., Jegga, A. G., Aronow, B. J., Lee, C.-K., Bernard, A., et al. (2015). Canonical genetic signatures of the adult human brain. *Nature neuroscience*, 18(12):1832.
[19] Josh Huang, Z. and Zeng, H. (2013). Genetic approaches to neural circuits in the mouse. *Annual review of neuroscience*, 36:183–215.
[20] Lein, E. S., Hawrylycz, M. J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A. F., Boguski, M. S., Brockway, K. S., Byrnes, E. J., et al. (2007). Genome-wide atlas of gene expression in the adult mouse brain. *Nature*, 445(7124):168–176.
[21] Menashe, I., Grange, P., Larsen, E. C., Banerjee-Basu, S., and Mitra, P. P. (2013). Co-expression profiling of autism genes in the mouse brain. *PLoS Comput Biol*, 9(7):e1003128.
[22] Ng, L., Bernard, A., Lau, C., Overly, C. C., Dong, H.-W., Kuan, C., Pathak, S., Sunkin, S. M., Dang, C., Bohland, J. W., et al. (2009). An anatomic gene expression atlas of the adult mouse brain. *Nature neuroscience*, 12(3):356–362.
[23] Oh, S. W., Harris, J. A., Ng, L., Winslow, B., Cain, N., Mihalas, S., Wang, Q., Lau, C., Kuan, L., Henry, A. M., et al. (2014). A mesoscale connectome of the mouse brain. *Nature*, 508(7495):207–214.
[24] Paus, T., Pesaresi, M., and French, L. (2014). White matter as a transport system. *Neuroscience*, 276:117–125.
[25] Robertson, M. M., Eapen, V., Singer, H. S., Martino, D., Scharf, J. M., Paschou, P., Roessner, V., Woods, D. W., Hariz, M., Mathews, C. A., et al. (2017). Gilles de la tourette syndrome. *Nature reviews Disease primers*, 3(1):1–20.
[26] Rybak, I., Shevtsova, N., Paton, J., Dick, T., John, W. S., Mörschel, M., and Dutschmann, M. (2004). Modeling the ponto-medullary respiratory network. *Respiratory physiology & neurobiology*, 143(2-3):307–319.
[27] Saggar, M., Sporns, O., Gonzalez-Castillo, J., Bandettini, P. A., Carlsson, G., Glover, G., and Reiss, A. L. (2018). Towards a new approach to reveal dynamical organization of the brain using topological data analysis. *Nature communications*, 9(1):1–14.
[28] Sizemore, A. E., Giusti, C., Kahn, A., Vettel, J. M., Betzel, R. F., and Bassett, D. S. (2018). Cliques and cavities in the human connectome. *Journal of computational neuroscience*, 44(1):115–145.
[29] Sporns, O. (2011). The human connectome: a complex network. *Annals of the New York Academy of Sciences*, 1224(1):109–125.
[30] Stafford, J. M., Jarrett, B. R., Miranda-Dominguez, O., Mills, B. D., Cain, N., Mihalas, S., Lahvis, G. P., Lattal, K. M., Mitchell, S. H., David, S. V., et al. (2014). Large-scale topology and the default mode network in the mouse connectome. *Proceedings of the National Academy of Sciences*, 111(52):18745–18750.
[31] Swanson, L. (2004). *Brain maps: structure of the rat brain*. Gulf Professional Publishing.
[32] Wasserman, L. (2018). Topological data analysis. *Annual Review of Statistics and Its Application*, 5:501–532.
[33] Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world'networks. *nature*, 393(6684):440–442.
[34] Zomorodian, A. and Carlsson, G. (2005). Computing persistent homology. *Discrete & Computational Geometry*, 33(2):249–274.