

Topological data analysis of *in vivo* Purkinje population calcium activity under different sensory stimulus conditions in awake mice

Tuan Pham

Abstract

In this project, I used topological data analysis (TDA) on Purkinje cell population activity during calcium imaging of awake mouse cerebellum under different stimulus conditions. I compared the topology of Purkinje network using variation of information as the distance metric with different null models and across different stimulus conditions. The results reveal low-dimensional geometric structures in the network and different stimulus modalities have interacting effects on the topology of the functional network under combinations of stimuli. However, using a multiperceptron (MLP) network reveals that using topology-based features do not have significant meaningful gain in decoding stimulus combinations (categories) in addition to using activity-based features.

Introduction

With the incredible advancement of high resolution neural recording, manipulation technology and computing power, the amount of data for both experimental and simulated neural population activity is increasing tremendously and can be quite overwhelming, for both analysis, interpretation and theory construction. Computational and analytical tools to interpret and build models from such volumes of data for hypothesis testing and exploratory purposes are catching up. Tools like topological data analysis (TDA) and network analysis inevitably are necessary for extracting and exploring potential structures and patterns within these data. Over the last two decades, there have been multiple applications of algebraic topology tools in neuroscience, extending analysis from traditional graph theory and network science methods. Aside from the more theoretical endeavors, these applications peruse TDA in neural population data across different subfields of computational neuroscience, most of which involve clique topology analysis. Examples span across different data domains and areas, from electrophysiological recordings in a specific area [1] (rodent hippocampal place cells), to brain-wide structural human connectome [2], as well as within detailed biophysical neural simulations and models [3]. Although there have been quite many population studies across other areas like neocortex, subcortical areas and hippocampus, not many network studies are done in the cerebellar Purkinje population recordings, even though these cells are among the popular neurons usually shown in textbooks with their beautiful complicated morphology, responsible for many developmental and learning functions such as motor learning. To my knowledge, there are no studies to date applying TDA in the cerebellum population data.

Hence, I want to apply TDA on Purkinje cell population and drawing comparison with different models of random graphs, similar to [1]. More specifically, I based my project around the ideas and methods from [1], which shows the geometric organization in hippocampal correlation structure across different states of the animal, including during wheel-running behavior and sleep. This was accomplished by comparing the Betti number curves by analyzing the homology of order complex between the data and those from generated random geometric networks, as well as comparing with a null based on shuffled data. In addition to comparison between the population and different null models, including geometric and block models, because the data are associated with different brief evoked sensory stimulus conditions, I explored how these different sensory modalities affect the topology features of the network. Moreover, I examined whether it is possible to classify the stimulus categories either from the topology features alone, or with combination of activity data.

Methods

1. Preprocessing

Data: The data are obtained from Dr. Christian Hansel's lab, with experiments designed and data collected by Silas Busch and Dr. Ting-Feng Lin. Briefly, the data are calcium imaging data (i.e. proxy for neural activity) of Purkinje population in awake mice, responding to brief sensory stimuli. Each trial is roughly 20 seconds of 60 Hz sampling rate, and the stimulus (if present) is briefly on at 10-second (see Fig. 1a). The raw fluorescence traces were median smoothed, normalized and further smoothed to obtain the classical normalized fluorescence traces dF/F_0 using matlab scripts. There are 78 regions of interests, assumed to be 78 distinct Purkinje neurons, in this dataset. Hence, throughout the

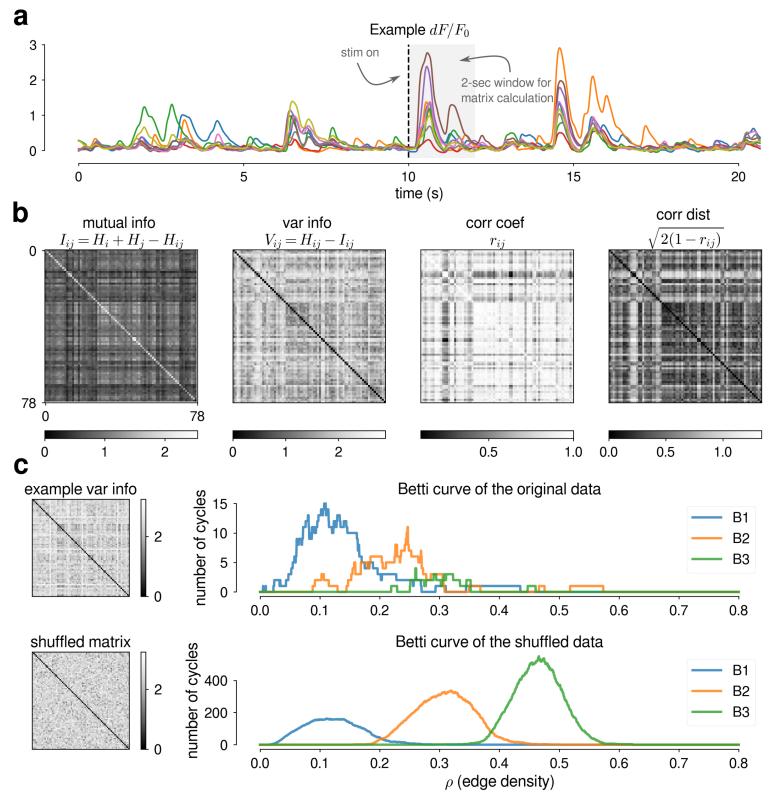


Figure 1: Data distance matrix and Betti curves. (a) Example activity traces. (b) Example similarity and distance matrices using information theoretic measures or correlation. (c) Example Betti curves as a function of edge density of an original variation information matrix (top) and a shuffled matrix (bottom)

project, I assumed $N = 78$ vertices for the null models.

Stimulus: There are 3 distinct stimuli modalities: L (light, i.e. visual stimuli), S (sound, i.e. auditory stimuli), P (air puff to paw, i.e. somatosensory stimuli). There are in total 8 different stimulus conditions, each consisting of simultaneous combinations of the different modalities: C (control, i.e. no stimuli), L, S, P, LS, LP, SP, LPS. Each of the 8 stimulus categories consists of 15-20 trials.

Similarity/distance matrix: Pairwise similarity or distance metric between different neurons are constructed based on dF/F_0 of a 2-second window after the stimulus onset (i.e. 10-12 second, see **Fig. 1a**). Similarity measures are either correlation coefficient r_{ij} between the windowed activity traces or mutual information $I_{ij} = H(X_i) + H(X_j) - H(X_i, X_j)$, where $H(A)$ denotes the entropy of A . Distance metrics are either correlation distances $\sqrt{2(1 - r_{i,j})}$ or variation of information $V_{ij} = H(X_i, X_j) - I_{ij}$. Both of these metrics do satisfy the triangle inequality. Examples of these different measures are shown in **Fig. 1b**. For the rest of the report, I will only be focusing on analysis done on the variation of information V_{ij} 's (`varinfo`) as I observed that correlation-related measures might be overestimating the strength of connectivity between these neurons when analyzed for only a short window. The resulting distance matrices are all symmetric.

2. Different null models

Shuffled data: Shuffled matrices are built by shuffling the $N(N-1)/2$ off-diagonal elements while still keeping symmetry (see **Fig. 1c** and [1]). As stated in [1], a shuffled version is equivalent to a type of noise model where entries of the matrix are independent drawn from a uniform distribution (IID models like in [1], [4]). For each trial of the original data matrix, I perform 3 shuffles.

Geometric nulls: Following [1], random geometric models are built by sampling $N = 78$ points from a hypercube of dimension $d \leq N$ (i.e. uniformly from $\mathcal{U}[0, 1]^d$). The corresponding distance matrices (Euclidean L2 norm, see **Fig. 2a**) could then be used to calculate homology. For dimension d , I generate 30 geometric null models.

Block nulls: Block models are constructed, taken inspiration from [4], to compare with the Purkinje population topology. Four different block configurations (30 models each) are used: `assort` (assortative), `disassort` (disassortative), `core` and `discore`. The similarity matrices W are built so that the “highly-connected” weight blocks are sampled from $\mathcal{N}(1.6, 0.5^2)$, and “lowly-connected” blocks are from $\mathcal{N}(1, 0.5^2)$. The “difference” matrix can be then constructed by $D = \max(W) - W$. Examples of these configurations are shown in **Fig. 2**.

3. TDA

Persistent homology: Persistent homology is computed with Rips complex using difference or distance matrices. In order to compare between the data and the different null models, I perform conversion of the distance parameter to edge density ρ of the graphs, and following [1], I only compare B_1, B_2, B_3 (to avoid confusion with entropy symbol, I use B to denote homology, Betti number related analysis instead of the conventional H symbol; see **Fig. 1c**, **Fig. 2a**, **Fig. 3a,b**). Additionally, I also analyze the “unconverted” version in B_0, B_1, B_2, B_3 (**Fig. 5a**). I use `gudhi` for TDA.

Betti curves: Following [1], [4], I compute the Betti curve per each homology dimension per each trial or null model, which is the Betti number as a function of edge density or distance (see **Fig. 1c**, **Fig. 5a** for examples). I denote this as β to later use for classification, with sub-sampling and smoothing.

Pairwise bottleneck distance: For each homology dimension, the mean bottleneck distance between the data stimulus categories or null models is computed (see **Fig. 3c₁**, **Fig. 5b₁**). Taken into account these pairwise bottleneck distance matrices, I visualize the relationship between the different stimulus categories and/or null models using graphs (see **Fig. 3c₂**, **Fig. 5b₂**). For comparison with null models, I only limit visualization to geometric models of dimension ≤ 10 as all the other models are too “far away” from the data for any meaningful comparison.

Persistent scores and topology features: Different features and scores are extracted from the bar codes $\{(b_i, d_i)\}$ for each homology dimension. I denote the set of these as ϕ to use later for classification.

- *Integrated Betti number*: the area under the Betti curves, using the trapezoid method
- *Mean, median, max lifetimes*: statistics of bar code lifetimes (i.e. $|d_i - b_i|$)
- *Persistent entropy*: entropy of the normalized lifetimes (see [5])
- *Mean algebraic pq of bar codes*: $\langle (d_i - b_i)^p (b_i + d_i)^q \rangle$ where $(p, q) \in \{(1, 1), (1, 2), (2, 1), (1, 3), (3, 1)\}$

4. MLP training for stimulus classification

Overview: A multi-perceptron (MLP) network with one hidden layer (50 units) is trained with inputs from activity of topology features to classify the 8 different stimulus categories. For simplicity, I consider the 8 different categories as distinct. However, future works need to consider multi-label classification instead, as the 8 different categories are combinations of 3 distinct modalities. Training is done with pytorch with google-colab resources.

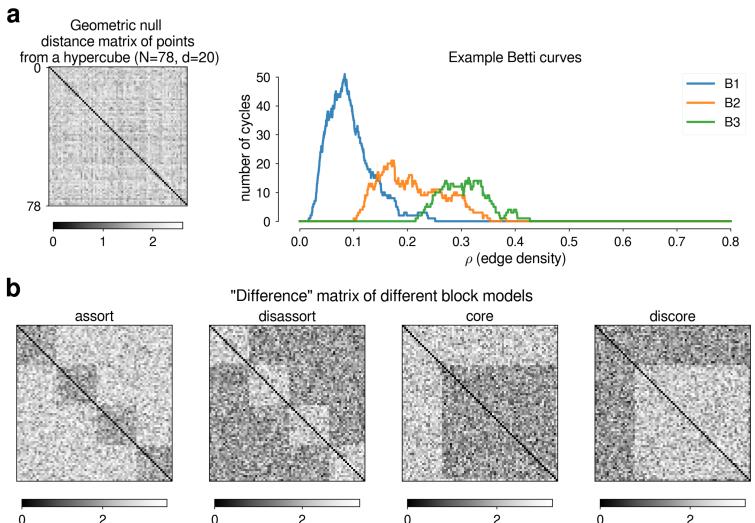


Figure 2: Null models. (a) Example geometric null model distance matrix and the corresponding Betti curves. (b) Example “difference” matrices of different types of block null models

The similarity matrices W are built so that the “highly-connected” weight blocks are sampled from $\mathcal{N}(1.6, 0.5^2)$, and “lowly-connected” blocks are from $\mathcal{N}(1, 0.5^2)$. The “difference” matrix can be then constructed by $D = \max(W) - W$.

Examples of these configurations are shown in **Fig. 2**.

Inputs: are combinations from 3 different sources:

- Average activity difference Δ between before and after stimulus onset. Specifically, denote $dF_i(a, b) = \frac{1}{b-a} \int_a^b \frac{dF_i}{dt}(t) dt$ with $a < b$, i.e. mean activity in $t \in [a, b]$ of the i -th neuron. Then $\Delta = \{dF_i(10, 12) - dF_i(5, 10)\}$ where $i \in [1, N]$.
- Persistent scores/features ϕ from persistent homology analysis with actual distance (instead of edge density).
- Betti curves β from persistent homology analysis with actual distance. These curves are sub-sampled and moving-average-smoothed. The Betti curves from 4 dimensions are concatenated into 1 vector to use as input.

Network: The activation function is tanh. The network is either vanilla mlp or with an optional batch-normalization layer after the input layer but before activation bn. The loss function is cross entropy loss (in the implementation, it is actually a negative log-likelihood loss with log_softmax applied at the output layer)

Training: is done in batches of size 5, for 12 epochs. Training is done either with stochastic gradient descent (SGD) with learning rate as 0.05 or Adam with learning rate as 0.001.

Classification and held-out: The task is classification of the stimulus categories. Since there are limited trials per category, I randomly choose one trial per category as the test set while the rest are used for training. And I evaluate the mean accuracy after doing that for 100 times. The data in **Table. 1** are either (mean \pm 95% confidence interval = 1.96 SEM) or the maximum accuracy at the end of training across the different train-test splitting instantiations.

Results

1. Comparison with null models

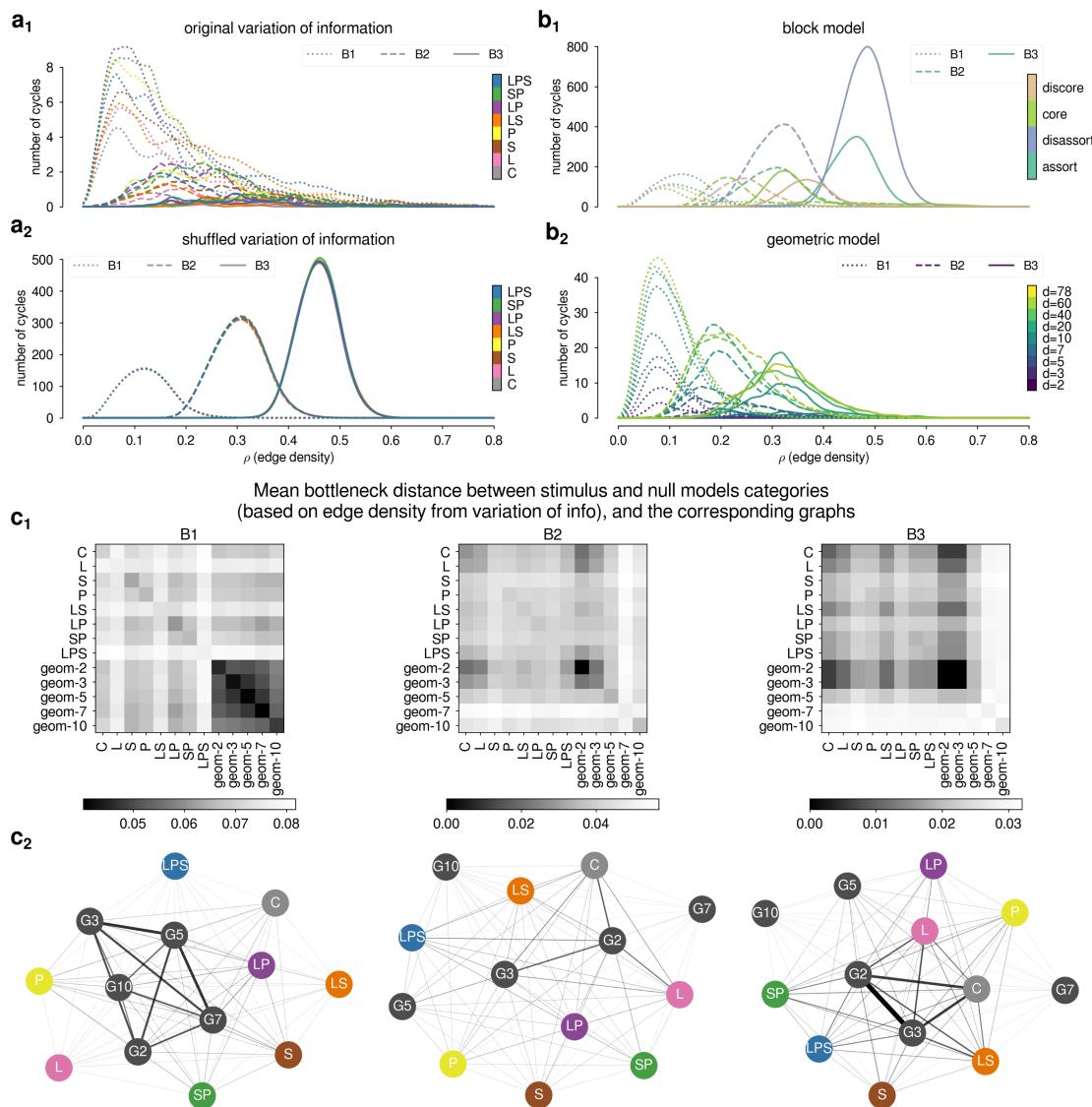


Figure 3: Comparison of Betti curves and bottleneck distances between Purkinje population topology and different null models. (a) Mean Betti curves as a function of edge density of the original data (a₁) and the shuffled (a₂) variation of information matrix data. Colors are different stimulus categories. (b) Mean Betti curves of the block null models (b₁, colors represent different block configurations) and the geometric models (b₂, colors represent different dimensions). (c) Mean bottleneck distance matrix (c₁) between different stimulus categories and a select few geometric null models; and the graphs (c₂) constructed from converting these distance matrix to similarity matrix (i.e. thickness of edge means smaller bottleneck distance).

With the Betti curves as a function of edge density, I can compare the topology of the original data with the different null models, as well as with the shuffled versions of the data distance matrices representing an IID null.

First off, the original Purkinje cell population topology is different from the IID null, via comparison of the Betti curves (**Fig. 3a_{1,2}**) and the integrated Betti numbers (**Fig. 4a₁**), suggesting that there exists some meaningful topology in the data and not completely random. This is similar to the observations made in hippocampal data [1].

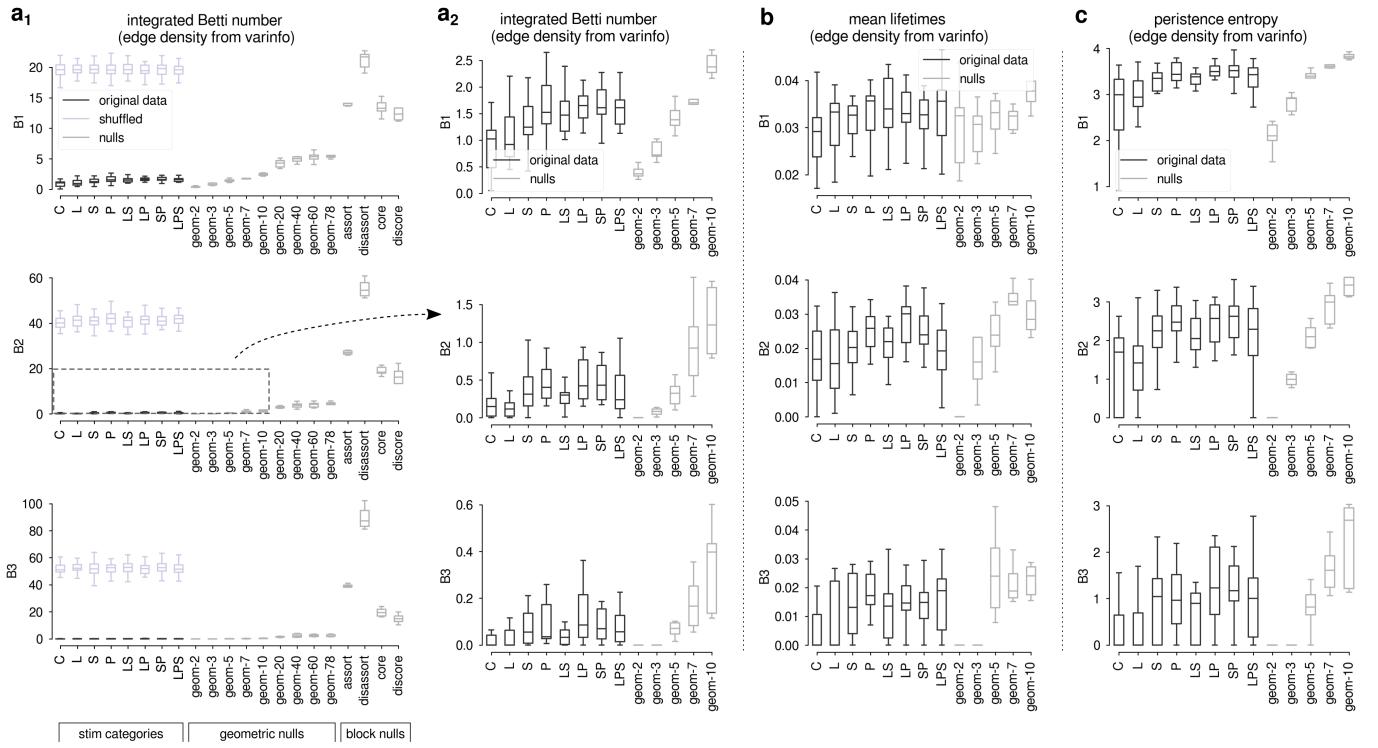


Figure 4: Comparison of persistent scores and features between Purkinje population topology and different null models. (a) Integrated Betti number from data, shuffled data and both geometric as well as block null models (a₁), and closer inspection of select few geometric models with dimension ≤ 10 (a₂). Similarly, (b) shows mean lifetimes and (c) shows persistent entropy.

Secondly, the topology of the population does not seem to resemble any of the block models as the number of cycles of these models are quite large and do not decrease with homology dimensions compared to original data (compare **Fig. 3a₁** and **b₁**). Hence, the integrated Betti numbers for these block models across dimensions would tend to be much larger than those of the original data. It is possible that if such block topology exists, it is much smaller. A way to test this in the future would be to consider more blocks or smaller number of elements per block.

Thirdly, when comparing with geometric models like done in [1], the analysis reveals some similarity between the Purkinje information network topology and geometric topology of low dimensions. This is supported with the overall decreasing number of cycles with increasing homology dimensions (compare **Fig. 3a₁** and **b₂**). However, I observe that the similarity, if it exists, would only exist for geometric models of dimensions less than 10. This is supported by inspecting the integrated Betti number, mean lifetimes and persistent entropy in **Fig. 4**. This is in contrast to the topology of the hippocampal place cell topology observed in [1], where similarity is observed in higher geometric model dimensions as well.

Additionally, upon closer inspection between the different stimulus categories and low-dimensional geometric models using the bottleneck distances (**Fig. 3c**), trials associated with C or L are more similar to lower dimensions geometric models in **B₂** (also somewhat recapitulated in **Fig. 4a₂** middle panel), while all of them are closer to $\text{geom} \leq 5$ in **B₃**.

In summary, the analysis shows that there possibly exists geometric topology of low dimensions (possible 2-10) in the Purkinje population either at spontaneous condition (no stimuli) or with stimulus. It is unclear whether block topology exists in the data at this point. Future endeavors require inspection of smaller blocks, as well other possible null models like a few mentioned in [4].

2. Comparison across different stimulus categories

I want to compare across different stimulus categories to see whether topology features could be used to decode out the different sensory conditions the animal was in. Inspecting using edge density does not reveal much “structure” in the bottleneck distance matrices (**Fig. 3c**). Additionally, it is possible that comparison using the actual distance is better: for example, a control (C) trial might reveal the same topology but with lower connectivity strength than a stimulus trial, and comparing the Betti curves as a function of the distance parameter without conversion to edge density might be more insightful for decoding. Hence, I repeated the analysis and extended it to also B_0 .

Inspecting the mean Betti curves (**Fig. 5a**) reveals that control (C) and light (L) trials generally evoke fewer cycles than other stimulus conditions (but C is still < than L), in all dimensions as the curves for the former two reach maximum and decrease earlier than the latter. Sound (S) and puff (P) trials seem to create more cycles generally, either with higher Betti curve peaks or wider curves. The combinations of stimulus modalities have interesting effects on the Betti curves in **B₁**, **B₂** (certain trends also follow in **B₃** but much noisier). For example, the presence of L brings down or shifts left the

curve when combined with either S or P. Additionally, the presence of P is usually accompanied with high Betti curve peaks. When all three modalities are present, the results are much more mixed.

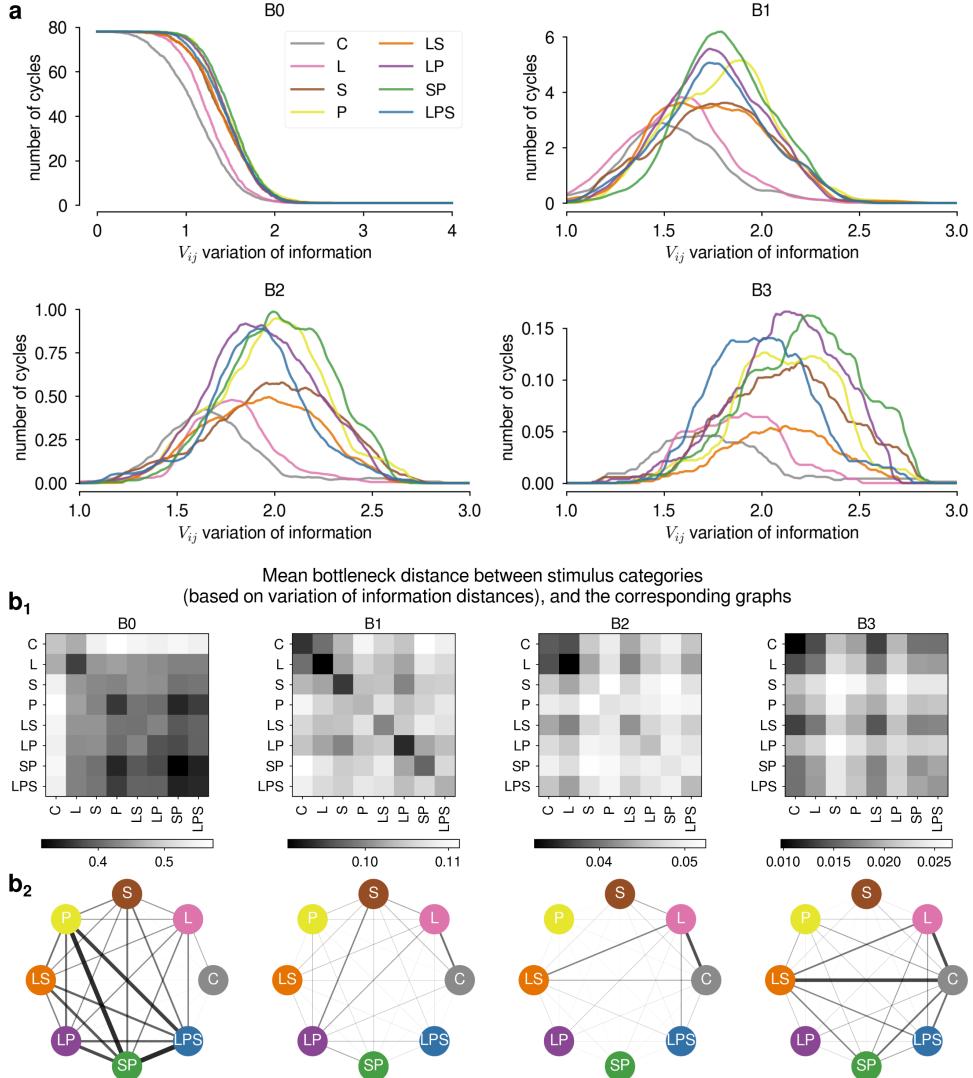


Figure 5: Comparison of Betti curves and bottleneck distances between Purkinje population topology of different stimulus conditions as a function of variation of information. (a) Mean Betti curves as a function of V_{ij} distance metric of the original data from variation of information matrix data. Colors are different stimulus categories. Different panels show different homology dimensions. (b) Mean bottleneck distance matrix (b₁) between different stimulus categories; and the graphs (b₂) constructed from converting these distance matrix to similarity matrix (i.e. thickness of edge means smaller bottleneck distance).

Roughly similar trends appear under inspection of the persistent scores (Fig. 6), though it is hard to say with significance due to low number of trials per category. In B_0, B_1, B_2 , integrated Betti number, mean lifetimes and especially persistent entropy generally reveal higher median values than control (C), and sometimes also than light (L). In B_2 , the effects of P presence is revealed as the four measures are generally higher when the categories contain P (i.e. P, LP, SP, LPS). It is harder to observe trends in B_3 for persistent scores. In all of the homology dimensions, the high variability in C might mean there were false positive trials, i.e. some irrelevant responses in the population activity to environment factors or maybe self-movements. This is difficult to assess without behavioral data. Higher number of trials might have blurred that out.

Upon inspection of the bottleneck distances (graphs) between these stimulus categories (Fig. 5b₂), it is clear that C is further from the rest in B_0, B_1, B_2 but still closer to L. P is much closer to SP and LPS than the rest especially in B_0 , but not necessary in other dimensions. In retrospect, a normalization to identity elements in the bottleneck distance matrices might have revealed better structure and relation between these categories. The fact that some of the categories have high mean bottleneck distances might mean that there is higher variability in the persistent diagrams.

Regardless, these comparisons show that the Betti curves and the persistent features might be useful in recognizing some categories or some modalities, for example at least whether a sensory stimulus is present in a trial versus a control trial. This prompts me to try using these features for classification in the following experiments.

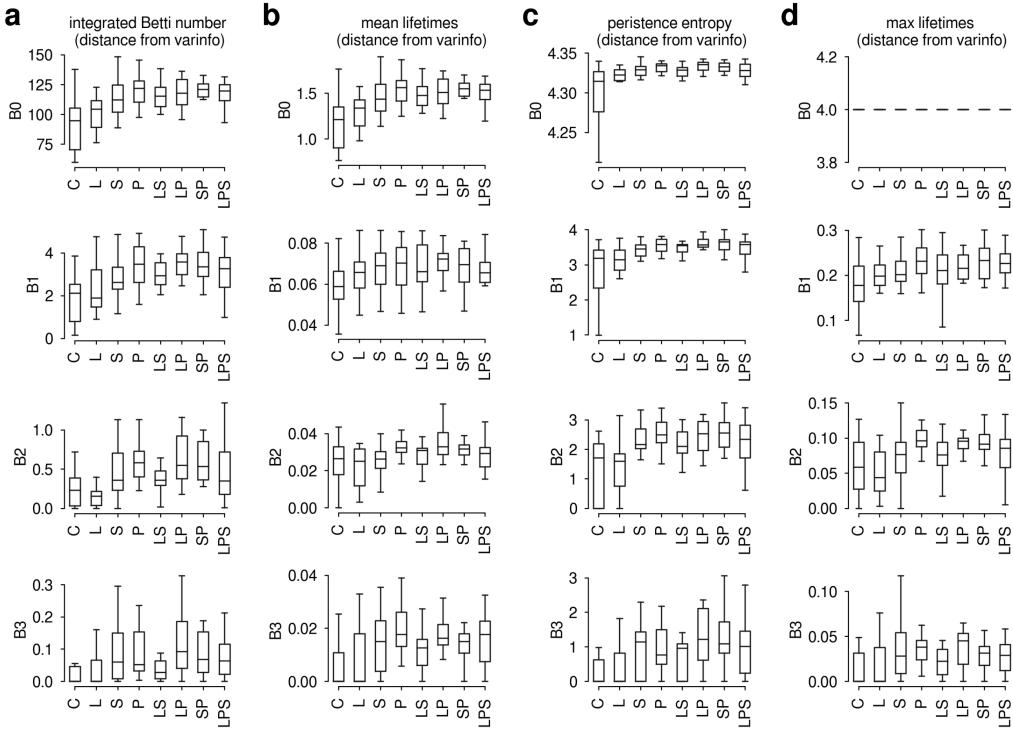


Figure 6: Comparison of persistent scores and features between Purkinje population topology across different stimulus conditions with distances. Integrated Betti number, mean lifetimes, persistent entropy and max lifetimes from variation of information data across different homology dimensions and across different stimulus conditions, with Betti curves analyzed as a function of the distance metric instead of edge density.

3. Classification of the different stimulus categories

For simplicity, I trained a simple MLP network with a single hidden layer with an option for batch-norm layer, trained using two different optimizers (hence 4 different training configurations). The inputs are combinations from the topology analysis and mean activity difference (see Methods). The testing is done in a one-trial-per-category hold-out fashion for 100 times. The reason for including activity features is to establish a baseline performance in which the model can rely on very accessible, easily computable features based directly on the activity.

As a side note, future computational experiments should consider other models (e.g. generalized linear models) and multi-label classification goals (instead of assuming single-label classification).

net-optim	inputs constructed on on						
	Δ (avr activ diff)	ϕ (pers scores)	β (Betti curves)	Δ, ϕ	Δ, β	ϕ, β	Δ, ϕ, β
mlp-SGD	24.12 ± 2.60	17.25 ± 2.24	17.38 ± 2.61	27.50 ± 2.84	22.12 ± 2.42	18.88 ± 2.36	26.50 ± 2.80
bn-SGD	26.50 ± 3.27	17.50 ± 2.62	17.75 ± 2.45	26.75 ± 3.00	21.50 ± 2.27	16.88 ± 2.33	21.50 ± 2.59
mlp-Adam	26.75 ± 2.59	18.00 ± 2.46	15.12 ± 2.36	27.25 ± 2.93	21.88 ± 2.84	15.12 ± 2.28	25.88 ± 2.38
bn-Adam	24.00 ± 2.91	19.88 ± 2.57	16.25 ± 2.09	25.62 ± 2.89	22.88 ± 2.48	15.50 ± 2.19	23.50 ± 2.80
[mean \pm CI ₉₅]							
mlp-SGD	62.50	50.00	50.00	62.50	62.50	50.00	62.50
bn-SGD	75.00	50.00	62.50	75.00	50.00	50.00	62.50
mlp-Adam	62.50	50.00	62.50	62.50	62.50	37.50	62.50
bn-Adam	62.50	75.00	50.00	75.00	50.00	50.00	50.00
[max]							

Table 1: Stimulus classification accuracies after training Left columns show the 4 different network configurations (network variants: vanilla vs with batch-norm layer; optimizer variants: SGD vs Adam). The inputs are constructed based on the combinations of 3 different sources: Δ (mean activity difference before and after stimulus onset), ϕ (persistent scores and features) and β (concatenated smoothed sub-sampled Betti curves). See Methods for details. The upper table shows the (mean \pm 95% confidence interval) of 100 instantiations of train-test splitting (one-trial-per-categories held out). The lower table shows the maximum accuracies achieved at the end of the training for each setup. Bold numbers represent the “best” in each column per table.

Results are shown in **Table 1**. The “best” accuracies across the 4 different training configurations **net-optim** (i.e. the bold ones per column in the upper table) show that the best mean performance is only a bit better than twice chance (chance is $1/8 = 12.5\%$). And it is sufficient to use just activity based features Δ to achieve this. Using persistent scores ϕ is moderately better than Betti curves β but either way is only better than chance by less than 8%. The only worthwhile combination comes from activity and persistent scores (Δ, ϕ) by roughly 1% across the different training configurations, but it might not be significant. Even if a rank significance test is performed, the performance gain is too small. Better feature building based on solely activity might already achieve better increase.

Inspecting the actual maximum achieved accuracies (lower half of **Table. 1**) reveals that there are some “best” cases where 6 out of 8 categories are correctly classified. And (Δ, ϕ) is not really much better than Δ alone. Inclusion of Betti curves apparent seems to degrade performance across different training configurations.

Discussion

In conclusion, topological analysis shows that, in this particular multisensory dataset, there might exist low dimensional geometric topology ($d \approx 2-10$) in Purkinje cell population at spontaneous and in response to sensory stimuli, by analyzing the variation of information distance metric of a short window after stimulus onset. However, future analysis should look at distance (or similarity) metric of the window after stimulus onset but also conditioned on the baseline window before the onset, e.g. partial correlation coefficient. Additionally, future analysis should consider block models of smaller block size and other null models.

Moreover, comparison between different stimulus categories shows that, at least when looking at the mean, different stimulus modalities affects the topology in a “mixed” way. In this particular dataset, somatosensory inputs (P) seem to induce more cycles and somewhat longer cycles than other modalities. It is to be seen whether this effect exists only in this area of recordings and in this individual mouse. It is possible that these effects are arbitrary, but might be sufficiently different for decoding purposes.

However, attempts at decoding the stimulus combinations using a simple MLP framework reveal that perusing features from TDA would only moderately increase performance when combined with activity-based features. Better performance might be gained by better construction of activity-based features alone without the need for TDA features. Future attempts should include multi-label classification goals and other simpler models like GLM, as well as regularization options during training. Additionally, features from persistent landscapes should also be considered.

Availability

Since these data are not published yet, I will not be able to make them public. The block and geometric null model data are in the repository, along with the codes for TDA, training and visualization on github at <https://github.com/tuanpham96/TDA-class-Winter2021/tree/main/proj>

Acknowledgement

The experimental data were obtained from Dr. Christian Hansel’s lab, the experiments were designed and data were collected by Silas Busch and Dr. Ting-Feng Lin. I would like to acknowledge and thank them for allowing me to use their data to play around for the project.

References

- [1] C. Giusti, E. Pastalkova, C. Curto, and V. Itskov, “Clique topology reveals intrinsic geometric structure in neural correlations,” en, *Proc. Natl. Acad. Sci. U. S. A.*, vol. 112, no. 44, pp. 13 455–13 460, Nov. 2015, ISSN: 0027-8424, 1091-6490. DOI: [10.1073/pnas.1506407112](https://doi.org/10.1073/pnas.1506407112).
- [2] A. E. Sizemore, C. Giusti, A. Kahn, J. M. Vettel, R. F. Betzel, and D. S. Bassett, “Cliques and cavities in the human connectome,” en, *J. Comput. Neurosci.*, vol. 44, no. 1, pp. 115–145, Feb. 2018, ISSN: 0929-5313, 1573-6873. DOI: [10.1007/s10827-017-0672-6](https://doi.org/10.1007/s10827-017-0672-6).
- [3] M. W. Reimann, M. Nolte, M. Scolamiero, K. Turner, R. Perin, G. Chindemi, P. Dłotko, R. Levi, K. Hess, and H. Markram, “Cliques of neurons bound into cavities provide a missing link between structure and function,” en, *Front. Comput. Neurosci.*, vol. 11, p. 48, Jun. 2017, ISSN: 1662-5188. DOI: [10.3389/fncom.2017.00048](https://doi.org/10.3389/fncom.2017.00048).
- [4] A. S. Blevins, J. Z. Kim, and D. S. Bassett, “Variability in higher order structure of noise added to weighted networks,” Jan. 2021. arXiv: [2101.03933 \[q-bio.QM\]](https://arxiv.org/abs/2101.03933).
- [5] A. Myers, E. Munch, and F. A. Khasawneh, “Persistent homology of complex networks for dynamic state detection,” en, *Phys Rev E*, vol. 100, no. 2-1, p. 022314, Aug. 2019, ISSN: 2470-0053, 2470-0045. DOI: [10.1103/PhysRevE.100.022314](https://doi.org/10.1103/PhysRevE.100.022314).