RESEARCH ARTICLE

Topology of synaptic connectivity constrains neuronal stimulus representation, predicting two complementary coding strategies

Michael W. Reimann [1‡]*, **Henri Riihimäki**[2], **Jason P. Smith**[2,3], **Jānis Lazovskis** [2,4], **Christoph Pokorny**[1], **Ran Levi**[2‡]

1 Blue Brain Project, École Polytechnique Fédérale de Lausanne (EPFL), Geneva, Switzerland, **2** University of Aberdeen, Aberdeen, United Kingdom, **3** Nottingham Trent University, Nottingham, United Kingdom, **4** University of Latvia, Rīga, Latvia

‡ MWR and RL are co-senior authors on this work.
* michael.reimann@epfl.ch

Abstract

In motor-related brain regions, movement intention has been successfully decoded from in-vivo spike train by isolating a lower-dimension manifold that the high-dimensional spiking activity is constrained to. The mechanism enforcing this constraint remains unclear, although it has been hypothesized to be implemented by the connectivity of the sampled neurons. We test this idea and explore the interactions between local synaptic connectivity and its ability to encode information in a lower dimensional manifold through simulations of a detailed microcircuit model with realistic sources of noise. We confirm that even in isolation such a model can encode the identity of different stimuli in a lower-dimensional space. We then demonstrate that the reliability of the encoding depends on the connectivity between the sampled neurons by specifically sampling populations whose connectivity maximizes certain topological metrics. Finally, we developed an alternative method for determining stimulus identity from the activity of neurons by combining their spike trains with their recurrent connectivity. We found that this method performs better for sampled groups of neurons that perform worse under the classical approach, predicting the possibility of two separate encoding strategies in a single microcircuit.

Introduction

Advances in experimental techniques have allowed us to record the *in-vivo* activity of hundreds of neurons simultaneously. This has gone along with new processing techniques to extract information from such spike trains. Some are based on the manifold hypothesis [1], which describes the high-dimensional spiking activity of large neuron populations as determined by a lower-dimensional space whose components may be aligned with behavioral (such as movement direction) or stimulus variables. The reconstruction of the underlying space can then improve the decoding of such variables.

The manifold hypothesis posits that activity is limited to the lower-dimensional space due to constraints imposed by the synaptic connectivity [1], however, details of this relation remain unclear. One identified constraint is that neurons with a large number of synaptic connections tend to spike at the same time as the population at large, while fewer connections allow neurons to spike when others are silent. Beyond this result, information on how the structure of synaptic connectivity shapes the neural manifold are scarce. This is in large part due to a lack of data: While a large number of neurons can be simultaneously recorded from, it remains challenging to determine the microstructure of their synaptic connectivity at the same time.

Conversely, in a model of a neural circuit, we have complete knowledge about connectivity as well as activity and can try to understand their relation. However, in order to study how connectivity shapes the structure of the neural manifold in biology, it is crucial that the connectivity of the model matches biology as closely as possible. Furthermore, to study this specific question, one additional requirement must be fulfilled: Encoding a lower-dimensional space in high-dimensional spike trains implies that the spiking activity is highly redundant. Such redundancy is often a means to overcome the presence of noise in a system [2]. Indeed, there are various sources of noise affecting neural activity and it has been demonstrated to be often unreliable [3–6]. Consequently, a model to study the structural implementation of a neural manifold needs to include these noise sources, to fully understand its function.

One such model is the rat neocortical microcircuit model of Blue Brain (*NMC-model*, [7]). It includes detailed synaptic connectivity that replicates a number of biologically characterized features [8, 9] and recreates a diverse set of experimentally characterized features of cortical activity [10–12]. It models morphological detail in the form of 55 morphological types of neurons and type-specific synaptic transmission between them [13]. Crucially, the model of synaptic transmission includes short-term plasticity [14–16] and both spontaneous release [17, 18] and synaptic failure [19], both important sources of noise in neural microcircuit [3, 20].

We first tested whether a lower-dimensional manifold could be used to describe the emergent activity of the model. To that end we used the model in a *classification task*: We subjected the model to a number of repeated synaptic stimulus patterns, recording the responses of all neurons. Then we used established techniques to reconstruct the neural manifold from the spiking activity and tested whether the values of its components were sufficient to classify the identity of the stimulus pattern injected at any given time.

Next, we investigated the dependence of classification accuracy on the connectivity of a neuronal population, thereby determining which local connectivity features give rise to a robust manifold capable of encoding information about the structure of a stimulus. We found significantly different numbers of motifs of three neurons in high-performing than in low-performing populations. To expand on this, we sampled neuronal sub-populations with uncommonly structured synaptic connectivity between themselves or to the rest of the population. Based on the results, a method to predict the classification accuracy of any sub-population based on a number of topological parameters was used. Finally, we developed an alternative method to reduce the dimensionality of spike trains that allowed us to reliably decode stimulus identity from the activity of populations where the classical technique failed.

Results

Detailed microcircuit model encodes stimuli in a lower-dimensional space

We began by investigating whether the manifold hypothesis, i.e. the hypothesis that spiking activity of a neuron depends on a linear combination of latent variables, can describe the results of simulations of the NMC-model. To that end we ran a simulation campaign subjecting 219,422 neurons (185,979 excitatory, 33,443 inhibitory) in the model to eight different
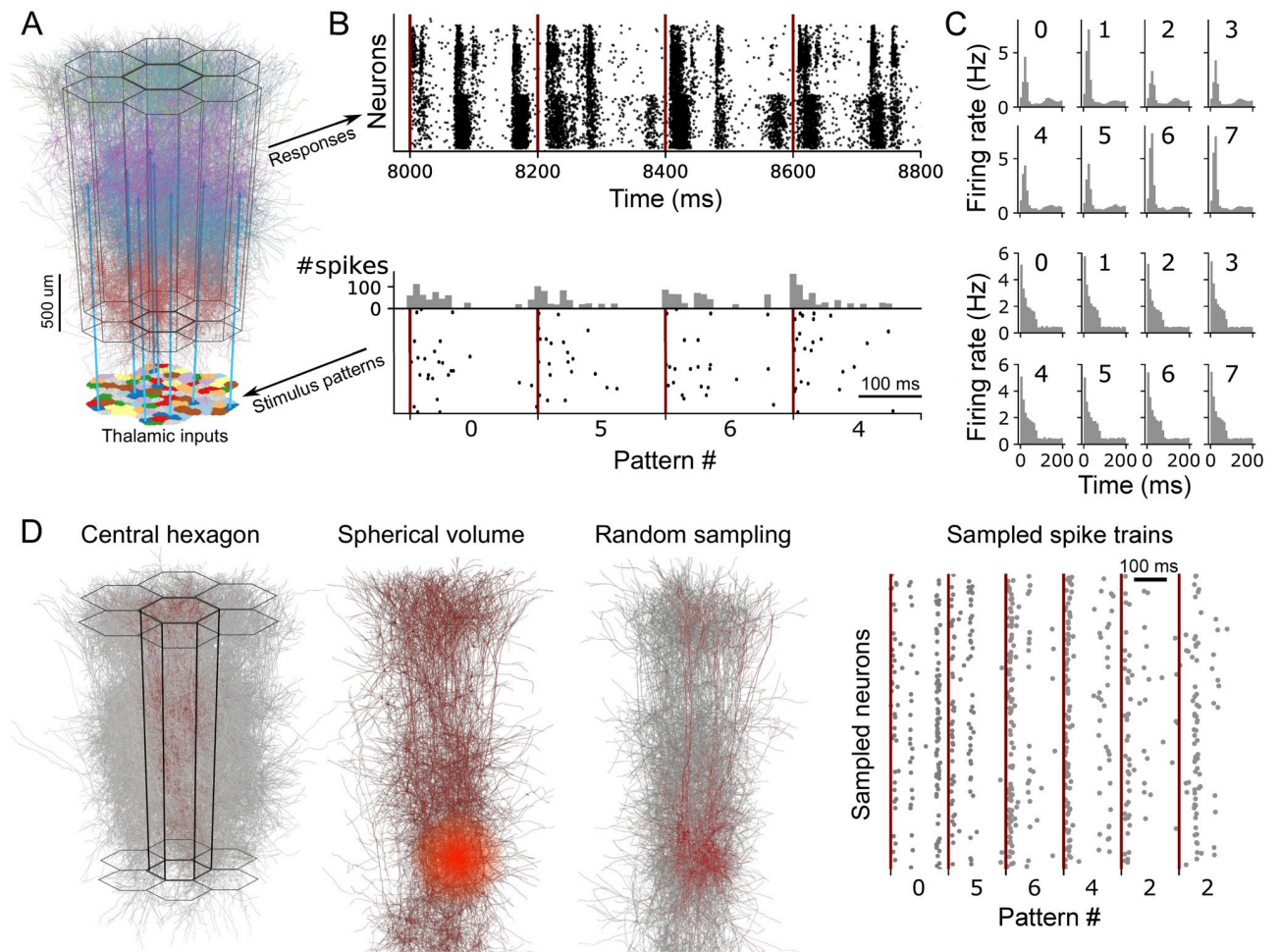
Fig 1. Synaptic input patterns applied to the detailed microcircuit model. A: The modeled microcircuit includes 2170 input fibers grouped into 100 bundles of nearby fibers modelling thalamo-cortical afferents. B: Spike trains are prescribed to the fibers to serve as stimulus patterns. Each pattern is associated with 10 randomly selected bundles of fibers. Each repetition of a pattern activates the associated fibers with an adapting Markov process, i.e. repetitions of a pattern use the same fibers, but with different stochastic spike trains. We apply a stream of 4495 repetitions of one of 8 stimulus patterns in random order (bottom) and simulate the microcircuits response (top). Raster plot of spiking responses is sorted by the layer of the simulated neuron, with layer 1 at the top and layer 6 at the bottom C: PSTHs of the input (bottom) and population spiking response (top) for each stimulus. D: Random sampling: We sample at random offsets by randomly picking 600 neurons from all neurons within a given radius and consider their spike trains for further analysis.

synaptic input patterns (Fig 1A and 1B). The stimuli were injected by activating thalamo-cortical afferent fibers that in turn activated synapses placed onto modeled dendrites according to a layer-specific density profile from the literature [21]. Each stimulus activated a random 10% subset of the fibers with an adapting, stochastic spiking process for 200 ms (see Methods). An adapting process leads to spike trains qualitatively matching the event-like structure [22] observed across thalamic systems, resulting from predominantly biphasic temporal filtering [22, 23] while requiring only few parameters. We chose a 200 ms inter-stimulus-interval, as we found that the population response to each stimulus decayed after a strong initial response back to a baseline firing rate within 100 ms (Fig 1C). Each stimulus was immediately followed by the next stimulus, randomly picked from the eight patterns, resulting in an uninterrupted, random stream. Each repetition of a pattern used the same synaptic input fibers, but different

randomly instantiated spike trains. In total, 4495 stimuli were presented, each pattern used 562 ± 4 (mean $\pm$ std) times.

Since connectivity in the model is distance-dependent, neurons near the edge of the modeled volume have fewer connections that would be filled in by neurons outside the volume. We restricted further analyses to the 31,346 most central neurons as they are least likely to be affected by this edge effect. Additionally, when analyzing how synaptic connectivity shapes the manifold of neuronal activity, the sign of a connection (excitatory or inhibitory) needs to be taken into account. To avoid this restriction on analysis methods, we analyzed only the excitatory population. Thus, all further analyses are applied only to the 26,567 most central excitatory neurons.

To approximate the recording of spike trains with extracellular electrodes, we sampled neurons within a given radius from points within the model and recorded their spike trains for further analysis (Fig 1D). The points were randomly picked from within a cylinder with a radius of $200\mu m$ and a height of $600\mu m$ placed at the geometric center of the model and oriented orthogonal to layer boundaries. Initially, we performed 25 such samplings within $175\mu m$. The number of samplings was set to conform with our second, structural sampling method presented in the next section. There, most of our sampling parameters display a pronounced change in their distributions around the value 25 [24].

We then analyzed the samples according to the manifold hypothesis by first extracting the hidden components through factor analysis (time bin size 10 ms, first 12 components extracted, see Methods). We then investigated whether the values of these components can be used to distinguish between the stimuli. Exploratory analysis revealed that within three of the strongest components different stimuli followed different average trajectories (Fig 2A), although individual trajectories were subject to a large amount of noise (S4 Fig).

Next, we wanted to understand which of the twelve extracted components contained information about the identity of the presented stimulus. The time course of a component during a single stimulus followed a simple trajectory of rapidly rising to a maximum (or negative minimum) value and decaying back to zero (Fig 2A). This allowed us to collapse the time course to the value furthest away from zero with minimal loss of information. We found that most of the time, this value differed significantly for the different stimulus patterns, with mean values for a given stimulus up to 1.6 standard deviations from the mean over all repetitions of all stimuli (Fig 2B). In other cases however, components never strayed more than half a standard deviation away from the overall mean.

We tried to quantify these differences by calculating for each component its expected usefulness for distinguishing between stimuli. To that end, we calculated the negative logarithm of the p-value of a Kruskal test against the null hypothesis that the mean value was identical for different stimuli (Fig 2C). This confirmed that pattern identity strongly affected the values for over half of the components in most volumetric samples. For some samples though, pattern identity affected only few and weaker components (Fig 2C, e.g. samples #8, #15, #19, #23). Finally, we performed a classification task attempting to discern stimulus pattern identity from the trajectories of the first 12 components. The linear classifiers were 6-times cross validated with a split into 60% training data and 40% validation data (see Methods). We found that the resulting classification accuracies partly reflected the differences between the samples. Where accuracies exceeded 80%, stimulus identity affected the trajectory of most components at very strong significance levels. Yet, the relation was more complicated in lower performing samples, where some samples classified better than others despite overall lower scores of its components (#16 vs. #17). This is likely because the information revealed by pairs of components can be partly redundant, which would not be captured by the test above.
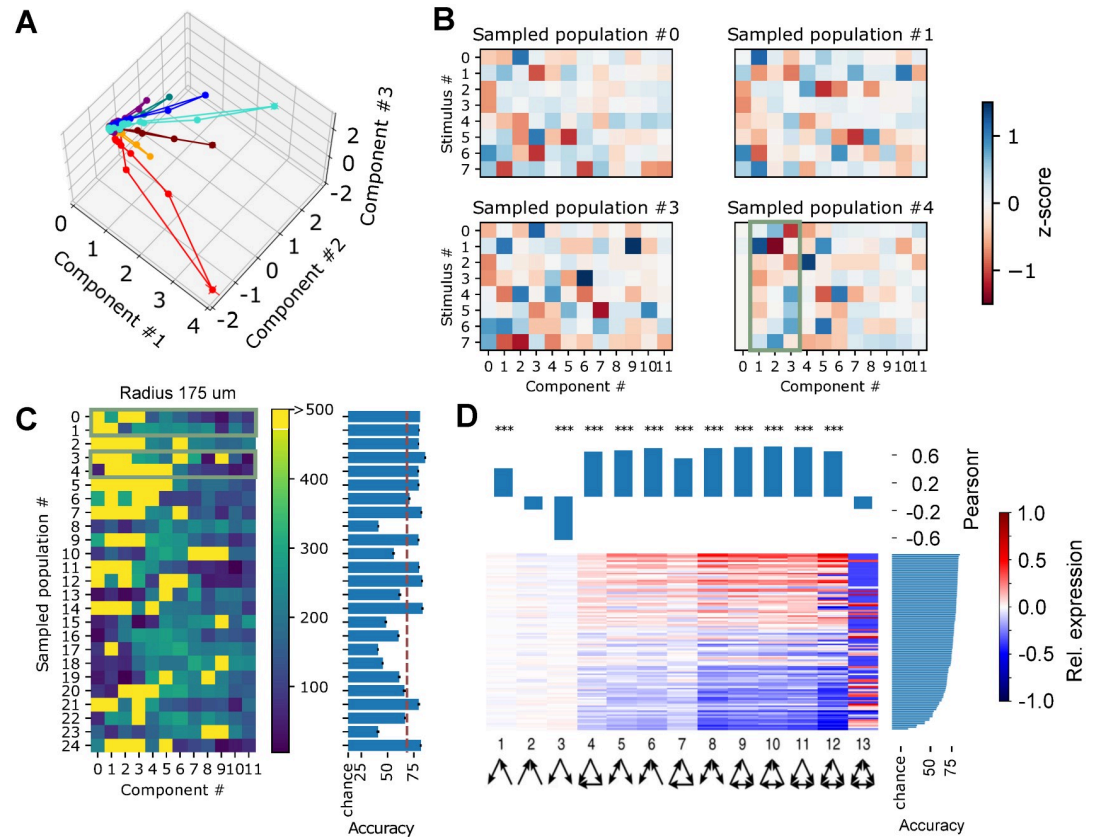
Fig 2. Analysis of volumetric samples according to the manifold hypothesis. A: Trajectories of three components for an exemplary volumetric sample with a radius of $175\mu m$. Responses to different stimuli are indicated in different colors (mean of 562 ± 4 (mean ± std) repetitions each). B: Mean of the most extreme value taken by individual components during a stimulus presentation, normalized as a z-score relative to the distribution over all trials and stimuli. Indicated for four random neuron populations, volumetrically sampled with a radius of $175\mu m$. Green outline indicates the three components shown in A. C: Left: For 25 volumetric samples with radius $175\mu m$: Negative decimal logarithm of the p-values of a test against the null hypothesis that their extreme values as in B have identical means (Kruskal test, n = 4495 stimulus repetitions). Green outlines indicate the samples shown in B. Right: Mean accuracy of a classifier of stimulus identity based on the trajectories of the 12 main components of the indicated volumetric samples (see Methods). D: Bottom: triad motif over- and underexpression in 125 volumetric samples with radii between 125 and 325 μm, relative to ER. Samples are sorted by classification accuracy (right). Top: Correlation between expression of a specific motif and classification accuracy in the samples (pearsonr).

We conclude that our simulations, combined with a volumetric sampling method lead to results that are comparable to the literature in the sense that different stimuli lead to different trajectories in a lower-dimensional manifold. However, the degree to which the manifold encodes stimulus identity varies between samples. According to the manifold hypothesis, the nature of the manifold is constrained by and therefore depends on the connectivity of the participating neurons [1]. We therefore aim to identify what feature of connectivity leads to more or less successful stimulus encoding.

We began by considering the presence of triad motifs in the volumetric samples. For each sample, we counted the occurrences of each possible motif of three connected neurons, and calculated their over- or under-expression compared to an Erdos-Renyi control model with the same number of neurons and connections (Fig 2D, bottom). When we compared the degree of expression of individual motifs to the classification accuracy, we found significant correlations for 11 out of 13 motifs. Specifically, for highly connected motifs with four or more

connections, over-expression was associated with higher accuracy and under-expression with lower accuracy. One apparent exception was the fully connected motif. However, we note that the actual and expected counts for this motif were extremely low, with 72% of samples having either zero or one instance. This difference of a single instance already marked the difference between strong under- or overexpression, making the analysis weak against noise and leading to a statistically non-significant result. While the result warrants further investigation, we cannot draw a conclusion on the effect of this motif yet. For weakly connected motifs with only two connections, this trend was largely reversed. As the degree of expression of motifs was normalized to the total number of connections in the sample, this indicates that strongly clustered connectivity leads to increased accuracy in the classification task. Unfortunately, the observed trend was very similar for all highly connected motifs, ruling out more fine-grained observations. Therefore, we explored a different approach to study the effect of connectivity.

A classification task in samples with uncommon connectivity reveals the significance of local graph structure

The hypothesis that differences of the connectivity between volumetric samples explain their different accuracies in the classification task leads to the following prediction: A neuron sample with an uncommon pattern of connectivity is more likely to depict an outlying classification accuracy—either very low or very high. To confirm this, we used a technique introduced in [24] to generate such samples with uncommon connectivity. We call this neuron sampling methodology *neuron-neighborhood* sampling and have recently validated it in a multiple measurements scheme [25].

Briefly, any of the 31,346 neurons in the model can be considered a *center* and its associated *neighborhood* contains in addition all neurons that are directly synaptically connected to it (Fig 3A). To find neighborhoods with uncommon connectivity, we defined 18 topological parameters that measure the structure of connectivity within, to, and from a neighborhood (see Table 1). We then sample the 25 neighborhoods containing at least 50 neurons and with the highest values for a parameter as the *champions* of that parameter.

The champions of a parameter were indeed outliers, with values that fell far outside the bulk of the distribution of the parameter over all neighborhoods (Fig 3B). To confirm that the connectivity of the champion samples deviated from the overall structure of the model, we first quantified the over- and underexpression of triad motifs in the volumetric samples (Fig 3C). We then compared the result to the expression of triad motifs in the champion samples and found large differences for all champions (Fig 3D). For each champion the prevalence of at least one motif was over five standard deviations from the mean of the volumetric samples.

However, the profiles of over- and underexpression were very similar for some pairs of parameters, indicating that their associated champions were redundant. We therefore removed champions and parameters with too strongly correlating triad profiles, reducing the number of parameters to 11 (see Methods, S3 Fig, Table 1).

We then repeated the stimulus pattern classification task, but this time using spike trains from the champion samples. As predicted, the resulting classification accuracy varied drastically between champions of different parameters, possibly reflecting their differently structured connectivity (Fig 4A). However, while the volumetric neuron samples always contained 600 neurons, the size of the champion samples depended on the connectivity of their centers and consequently varied. When we compared the size of the samples to the classification accuracies, we found a strong correlation (Fig 4B, grey diamonds and blue dots).

This led to two questions: First, is this observed difference in accuracy a result of different connectivity—in this case, a lower combined in- and out-degree of the center—or simply an
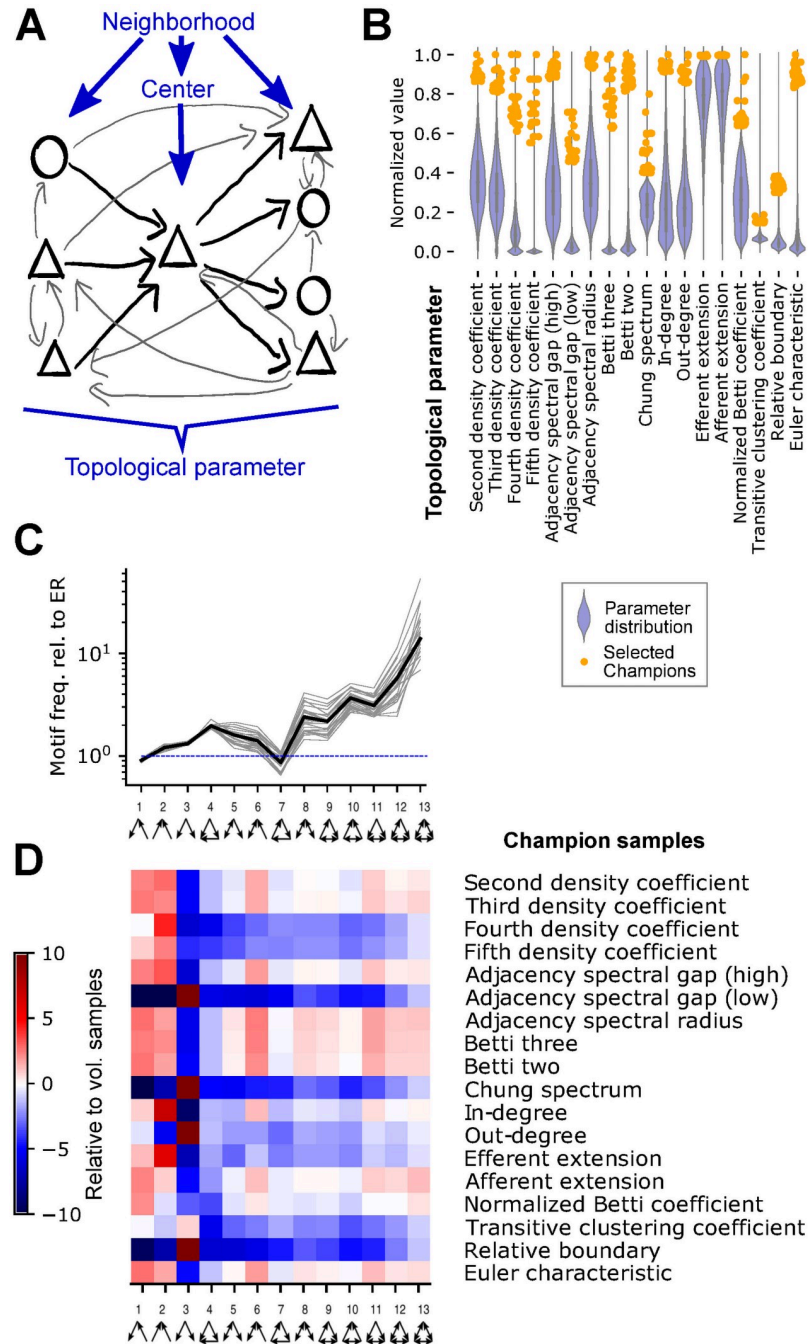
Fig 3. Neuron-neighborhood sampling with uncommon connectivity. A: Neuron-neighborhood sampling: Any neuron can be a *center*. Its associated *neighborhood* comprises the center, all synaptically connected neurons, and the connections between them. We then calculate various topological parameters of the neighborhood's connectivity. B: Blue violinplot: Distribution of various topological parameters, normalized between 0 and 1. Orange dots: Location of the *champions*, i.e. the 25 neighborhoods of at least 50 neurons with the highest values for a parameter. C: Frequency of triad motifs in volumetric samples with a 175μm radius relative to an Erdos-Renyi graph with the same number of nodes and edges. Grey: 25 volumetric samples; Black: their mean. D: Over- and under-expression of triad motifs in the connectivity of *champion* samples of various parameters, normalized with respect to mean/std of the volumetric samples.

Table 1. Topological parameters used.

Parameter	Shorthand	Explanation	Type
Adjacency spectral gap (high)		Difference between two largest moduli of eigenvalues of adjacency matrix	Recurrent
Adjacency spectral gap (low)	ASG	Lowest nonzero modulus of an eigenvalue of adjacency matrix	Recurrent
Adjacency spectral radius		Maximum modulus of an eigenvalue of adjacency matrix	Recurrent
Extension		Number of neurons connected to but not in neighborhood	
Afferent extension	AE	* for afferent connectivity	Afferent
Efferent extension	EE	* for efferent connectivity	Efferent
Betti numbers		Betti numbers of the neighborhoods	
Betti two		Second Betti number	Recurrent
Betti three		Third Betti number	Recurrent
Chung spectral gap		Smallest eigenvalue of largest strongly connected component of neighborhood	Recurrent
Euler characteristic	EC	Euler characteristic of the neighborhood	Recurrent
Density coefficients		Number of k-simplices, divided by number of k-1-simplices and normalized	
Second density coefficient		* for second dimension	Recurrent
Third density coefficient		* for third dimension	Recurrent
Fourth density coefficient	4DC	* for fourth dimension	Recurrent
Fifth density coefficient	5DC	* for fifth dimension	Recurrent
In-degree	ID	Total in-degree of the center	Afferent
Out-degree	OD	Total out-degree of the center	Efferent
Normalized Betti coefficient	NBC	Sum of normalized Betti numbers,	Recurrent
Relative boundary	RB	Number of connections into neighborhood, divided by connections within neighborhood	Afferent
Transitive clustering coefficient	TCC	Number of directed 3-cliques containing the center, divided by number of possible such cliques	Recurrent

Type denotes which connections determine the parameter: Only recurrent connections within the neighborhood, or additionally afferent connections into the neighborhood, or efferent connections. Parameters without a shorthand assigned were removed for being redundant. For details, see Methods.

https://doi.org/10.1371/journal.pone.0261702.t001

artifact of the analysis—i.e. a consequence of using fewer spike trains in the classification task? And second, do the various topological metrics capture relevant features of connectivity beyond simply the size of the sampled neighborhood?

To address the first question, we repeated the pattern classification task for the three largest champions at reduced sizes. That is, we used only 90%, 70%, 50%, 25% or 15% of the neurons in the champion neighborhoods of ED, ID and OD for classification (see Methods). The classification accuracies of these subsampled champions was only slightly reduced compared to the performance of the complete sample (Fig 4B, red markers). Importantly, the subsampled champions performed at the top end or better than complete champions of comparable size. This indicates that for complete champions of smaller size (such as relative boundary), the lower number of spike trains being analyzed does not fully explain their reduced classification performance. Instead, the smaller size indicates weaker connectivity of the center to the rest of the population, which may reduce the neighborhood's ability to successfully encode stimulus identity.

To address the second question, we first generated a large number of additional neighborhood samples. This time, we sampled 1276 neighborhoods by picking up to 25 centers randomly from each of the 55 morphological types of neurons in the model [7, 26]. We then subjected the samples to the stimulus pattern identification task (S5 Fig). This provided us with a large number of additional data points that were not biased towards extreme connectivity patterns, but represented typical connectivity of the model, while still providing variable classification accuracy. As before, classification accuracy in these samples depended on the
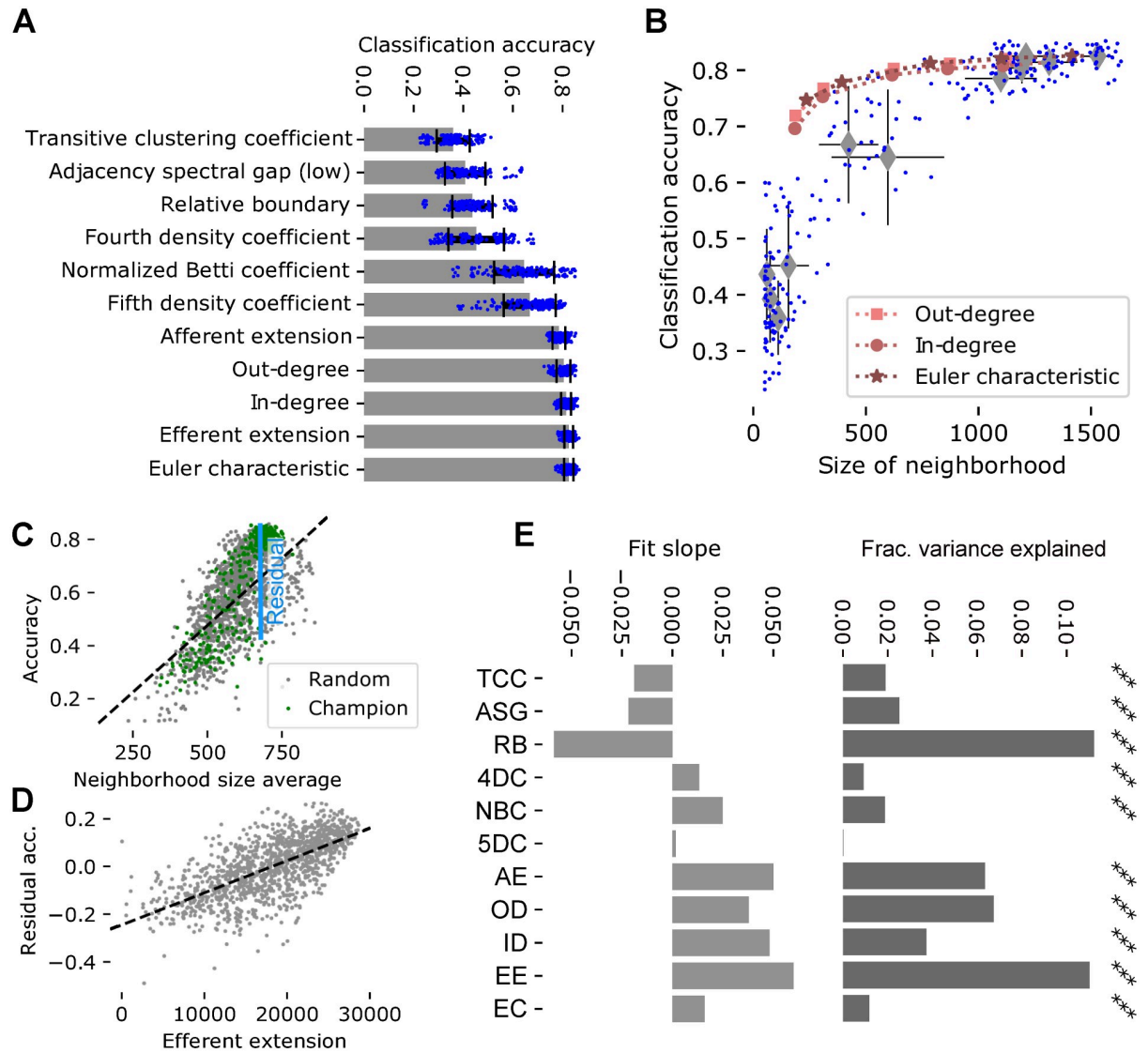
Fig 4. Classification accuracy in champion samples. A: Accuracy of a linear classifier to decode stimulus identity from the spike trains of the champion samples. Blue dots: Accuracies of 6-times cross validation for 25 champions (see Methods); bar and errorbars: mean and standard deviation. B: Size of champion neighborhoods against their decoding accuracy. Blue dots: individual neighborhoods; grey diamonds and errorbars: mean and standard deviation of champions of a parameter. Red markers: For the three largest classes of champions as in the legend, subsampled at 90%, 70%, 50%, 25% and 15%. C: Values of the neighborhood size average measurement (see Methods) for champion (green) and randomly picked (grey) neighborhoods against their classification accuracy. Black line: Linear fit to both; indicated in blue: residual of the fit. D: Efferent extension against the residual accuracy indicated in C for randomly picked neighborhoods. Grey dots: 1276 random neighborhoods; black line: linear fit. E: Left: Slopes of linear fits of residual accuracy against normalized values of the indicated parameter for randomly picked neighborhoods. Right: Fraction of variance of the residual accuracy that is explained by the linear models over a model that takes only the morphological type of the center into account (see Methods). Black stars indicate Bonferroni-corrected significance (***: $p < 0.0001$).

https://doi.org/10.1371/journal.pone.0261702.g004

sizes of neighborhoods, in particular we found a linear relation to the *neighborhood size average (NSA)*, i.e. the mean over all neurons in a sample of the size of the associated neighborhood (Fig 4C, grey dots). Corresponding data points for champion samples matched the fit, indicating that it captured the effect of neighborhood size we found earlier (Fig 4C, green dots). We performed a linear fit of the data (Fig 4C, black line) that allows us to predict the effect of NSA

on classification accuracy. Subtracting the prediction from the measured accuracies yields the *residual accuracy*, i.e. the accuracy with the effect of neighborhood sizes largely removed (Fig 4C, blue). As the NSA measure does not depend on the presence of a center in the sample, we will later be able to generalize this concept to the volumetric samples.

We studied the relation between values of the various topological parameters and the residual accuracy using data from the randomly sampled neighborhoods (for an example see Fig 4D). Specifically, for a given topological parameter, we pooled the data from all random neighborhoods and used them to fit a linear model of the combined effects of the parameter values and the morphological types of their centers. We then calculated the additional fraction of variance this model explained over one only taking the morphological type into account (Fig 4E, right). This additional step let us rule out explanations where connectivity patterns captured by the topological parameters are merely correlated with the morphological types and have no direct effect on residual accuracy. Additionally we considered the slope of the linear fit in each parameter to assess the direction of its effect. We found statistically significant effects for all parameters except the fifth density coefficient (t-test, $p < 0.0001$ Bonferroni-corrected; Fig 4E). This indicated that the topological parameters were capable of measuring features of connectivity conducive to encoding the identity of a stimulus pattern in a lower-dimensional space. The largest impact could be found for measures impacted by both recurrent connectivity within the sample and its connectivity to the rest of the population (RB, AE, EE). The positive impact of the AE measurement indicates that a sample requires strong external input, but the negative impact of RB shows that strong internal connectivity is even more important. Strong effects also of ID and ED reinforce a general trend of more synaptic connectivity increasing performance, but also a number of parameters measuring structure rather than quantity (ASG, EC, 4DC, NBC) had significant effects.

Predicting classification accuracy based on topological parameters in volumetric samples

Finally, we tried to generalize our ability to predict the classification accuracy from the topology of synaptic connectivity to the volumetric samples. As before, we calculated the mean of the sizes of neighborhoods associated with neurons contained in the samples, but this time for volumetric samples with radii between $125\mu m$ and $325\mu m$. Based on this, we predicted their classification accuracies using the linear fit from before, and subtracted the prediction to gain residual accuracies. We then calculated the values of the topological parameters based on the connectivity of the samples (Fig 5A, left) and analyzed their relation to the residual accuracies as in the previous section. However, the parameters in-degree, out-degree and transitive clustering coefficient were defined with respect to the center of a neighborhood (see Table 1) and consequently undefined for volumetric samples. We instead used the mean of in-degree, out-degree and clustering coefficient over all neurons in a sample. Most parameters calculated this way had no significant effect on the residual accuracies with only three exceptions (Fig 5B). It thus appears as if the ability of most topological parameters to capture significant connectivity patterns is linked to neighborhood structure of the samples we used before.

Therefore, we developed the following method to generate a *synthetic value* of a topological parameter for a volumetric sample that takes its neighborhood structure into account. First, calculate the relative size of the overlap between each neighborhood in the model and the sample. Next, find the N neighborhoods with the strongest overlap (Fig 5A, right for an example with $N = 3$). Note that the center of a neighborhood does not need to be contained in the sample for a strong overlap. Finally calculate the synthetic value as the weighted mean of parameter values of the N neighborhoods, where the weights are proportional to the relative overlap sizes.
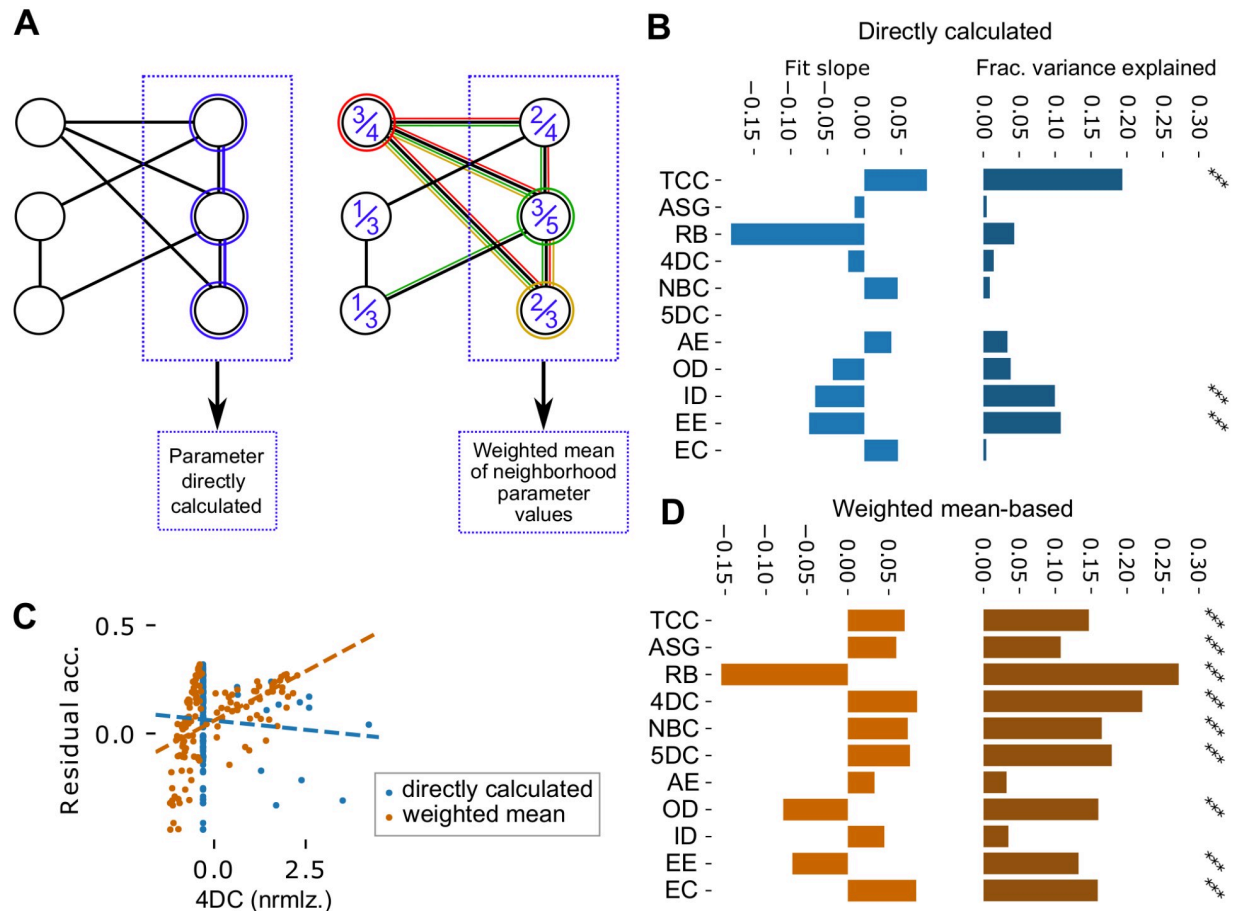
Fig 5. Topological parameters in volumetric samples. A: Two ways of calculating values for topological parameters in volumetric samples. Left: Values are calculated directly on the synaptic connectivity within the sample (blue neighborhood). Right: Generation of "synthetic" values of topological parameters that take the neighborhood structure of the sample into account: The size of the overlap of each possible neighborhood with the sample is calculated, then a weighted average of the parameter value for the $N = 3$ strongest overlapping neighborhoods is used (red, yellow, and green neighborhoods with relative overlaps of 3/4, 2/3, and 3/5 respectively). B: Left: Slopes of linear fits of residual accuracy against normalized parameter values that were directly calculated. Right: Fraction of variance of the residual accuracy that is explained by the linear models over a model that takes only the radius of the sample into account (see Methods). Black stars indicate Bonferroni-corrected significance (***: $p < 0.0001$). Shorthand parameter names see Table 1. C: Comparing directly calculated parameter values (blue) to weighted mean-based, "synthetic" values of the fourth density coefficient. D: As in B, but for synthetic values of topological parameters.

For each topological parameter, an optimal value for N was determined that maximized the correlation between the resulting synthetic values for volumetric samples and their residual classification accuracies (see Fig 5C for an exemplary result). Using this method, all but two investigated parameters provided a statistically significant effect (t-test; $p < 0.0001$ Bonferroni-corrected; Fig 5D; S7A Fig).

The optimal number of neighborhoods to include in the weighted mean (N above) tended to be low, under 300 neighborhoods for most parameters, with larger numbers leading to a gradual decline in correlation with classification accuracy (S7B Fig). Exceptions were TCC, EE and OD, where the optimal solutions were weighted means including all 31,346 neighborhoods. Curiously, this was accompanied by a switch of the sign of the correlation as more neighborhoods were included (compare Figs 4E to 5D). This indicates that at least these two classes of parameters capture different ways in which connectivity affects the performance of a neuron sample in the classification task.

In summary, we found that our method to predict the classification accuracy of a sample from synaptic connectivity can also be employed in the less artificial volumetric samples. A strong predictor for high accuracy is the presence of strongly interconnected motifs formed by a neuron and most of its neighborhood (small value of RB combined with large values of 4DC, 5DC, TCC). Incoming rather than outgoing connectivity is helpful (opposite effects of AE and ID vs. EE and OD).

An alternative decoding method for low-performing populations, based on graph topology

Having identified the connectivity patterns that improve a population's ability to encode stimulus patterns, the question remains what the role of the low-performing populations might be in the neural computations of a microcircuit. These are populations containing neurons forming small neighborhoods, that is, neurons with structurally weak coupling to the rest of the population. A dichotomy between weakly and strongly coupled neurons has been characterized before in the form of *soloists* that spike when the rest of the population is silent, and *choristers* that spike together with the rest [27]. This was characterized in terms of the *coupling coefficient* of a neuron, calculated as the correlation of its binned spike train with the firing rate of the rest of the neurons (see Methods). Neurons with a value below / above a control distribution calculated from shuffled spike trains are considered soloists / choristers. We therefore investigated whether the low-performing samples and neighborhoods contained a larger amount of soloists.

We began by calculating the coupling coefficient of all excitatory neurons in the population and comparing its distribution to two controls with shuffled spike trains: One that preserved only the overall firing rate of the entire population and one that preserved firing rates of individual neurons (see Methods, Fig 6A). We confirmed that in our simulation both soloists and choristers, that is, neurons with an unexpectedly low or high coupling coefficient emerged. Next, we investigated whether this property depended on the size of the neighborhood that a neuron is the center of (Fig 6B). We found a strong, positive correlation between these properties, albeit only for centers in layers 1 to 5. Curiously, for centers in layer 6 the relation reversed, with larger neighborhoods apparently reducing the coupling coefficient of their neighborhood. At the same time, neurons in layer 6 depicted the highest coupling coefficients.

Having found that the size of a neighborhood affects both classification accuracy and coupling coefficient, it was no surprise that we also found a strong correlation between the coupling coefficients of neurons in a neighborhood and its classification accuracy (Fig 6C). However, once again, neighborhoods with centers in layer 6 reversed this correlation. For the volumetric samples, we analyzed the effect of mean coupling coefficient on residual accuracy, finding the same strongly positive correlation as for layers 1 to 5 (Fig 6D). The coupling coefficient explained over 30% of the variance of residual accuracy, more than any of the topological parameters (compare inset to Fig 5D). Taken together, this confirmed that neurons with an uncommonly low coupling coefficient below 0.1 were more prevalent in neighborhoods and volumetric samples that did not perform well in the classification task.

In order to further explore the ways in which stimuli are encoded in microcircuit activity, we employed an alternative decoding method based on *topological featurization* (Fig 6E [24]). Similar to the manifold-based method, it begins by binning the spike trains of individual neurons in a neighborhood into 10-ms time bins. However, the way the dimensionality of this result is reduced, differs drastically. In each time bin, we consider all neuron members of a neighborhood that are spiking, and the connections between them. We can then calculate any
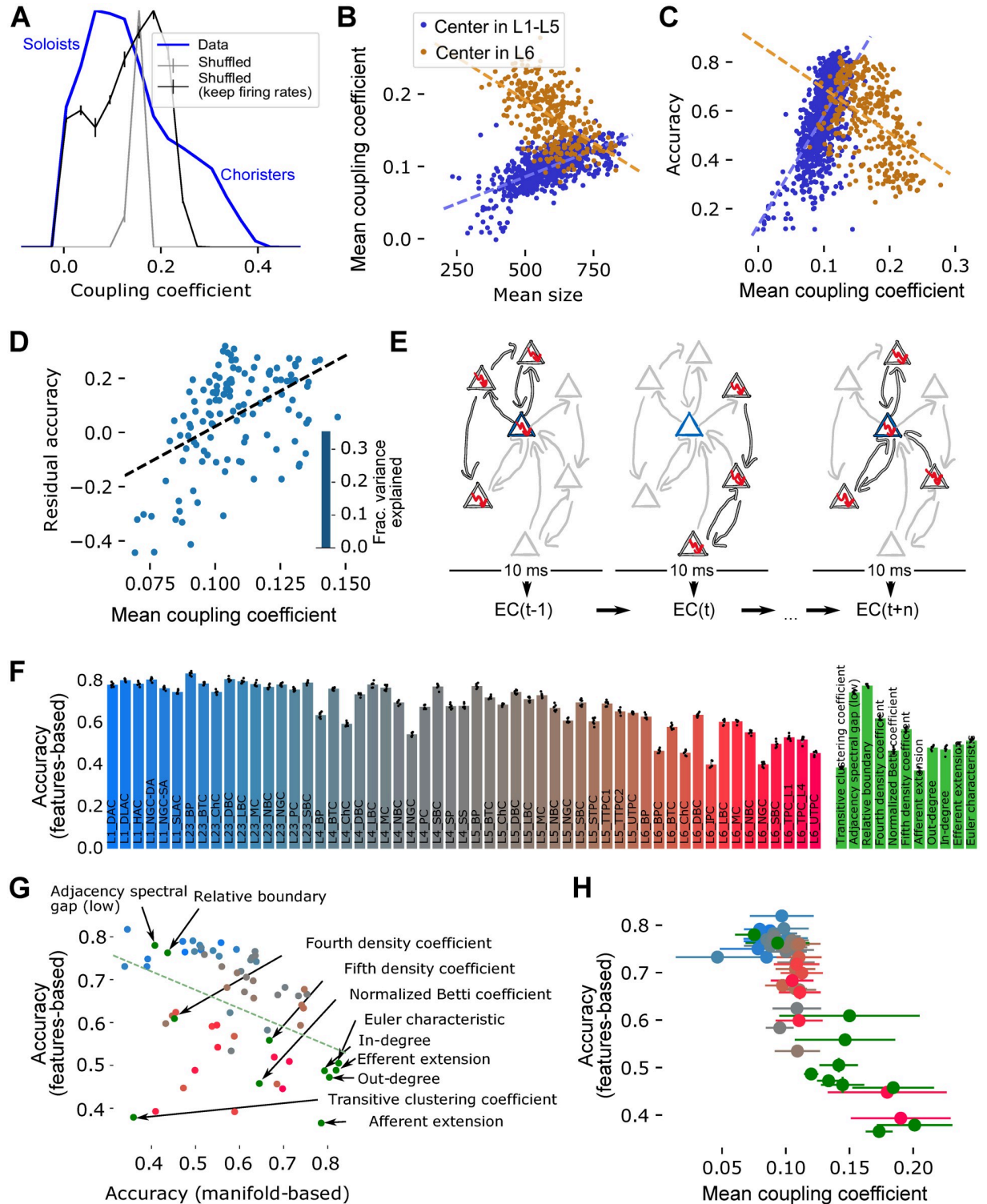
Fig 6. Coupling coefficient and topological featurization. A: Distribution of the *coupling coefficient*, as in Okun et al., 2015, of all excitatory neurons in the model. Blue: data; grey lines with errorbars: Shuffled controls that keep or shuffle mean firing rates of individual neurons. Neurons with an unexpected low coefficient are "soloists", a high coefficient indicates a "chorister". B: Neighborhood size against the coupling coefficient of the center for center neurons in layers 1–5 (blue) and in layer 6 (orange). Dashed lines: linear fits. C: As in B, but coupling coefficient against classification accuracy of the neighborhoods. D: Mean coupling coefficient of neurons in a volumetric sample against its residual classification accuracy, as in Fig 5C. Inset: Fraction of variance explained as in Fig 5D. E: Classification based on *topological*

featurization: For a given neighborhood, we consider in each 10 ms time step only the subset of neurons that are firing. We then calculate the value of the Euler characteristic (EC) of the graph of connectivity between them. We classify stimulus identity based on the EC time series of 25 centers. F: Left: Results of the classification for centers randomly picked from the various morphological types in the model. Errorbar: std of cross validation. Right: Same, for the various champions. G: Mean classification accuracies using the manifold-based method against accuracies based on topological featurization. Red/blue dots: Random centers as in F; green dots as labelled. H: Accuracies as in F against the mean coupling coefficient of neurons in the neighborhoods. Errorbars: std over coupling coefficients.

parameter of the graph of the spiking sub-network, yielding a time series of parameter values that represent the time course of activity of the neighborhood (see Methods). We used as parameter the Euler characteristic (*EC*) of the flag complex, which has been identified as one of the most promising in [24]. This can be thought of as a measure of the structure of synaptic connectivity between concurrently active neurons. As this reduced the dimensionality to a single dimension with twenty time steps for a neighborhood, we grouped 25 neighborhoods together and used their pooled time series in the same linear classification method as before. The neighborhoods pooled together were either champions of the same parameter, or their centers randomly picked from the same morphological type.

As for manifold-based classification, the results depended on the type of neighborhood. For neighborhoods based on picking centers randomly from a given morphological type, we found an overall gradient from accuracy around 80% for centers in superficial layers, to under 50% for centers in layer 6 (Fig 6F, left). For pooled champions, the accuracy varied comparably between 40% and 80%, depending on the topological parameter they were champion of (Fig 6F, right).

Astonishingly, the results were almost completely the inverse of the manifold-based classification: neighborhoods that performed well in the manifold performed poorly for topological featurization and vice versa (Fig 6G). Only some neighborhoods with centers in layer 6 and the champions of the transitive clustering coefficient performed equally badly for both methods. This indicated that topological featurization is a classification method more suitable for the samples that performed poorly in manifold-based classification, that is, samples with a large number of soloist-type neurons. Indeed, we found a strong negative correlation between the mean coupling coefficient of a neighborhood and its topological featurization-based classification accuracy (Fig 6H). We therefore predict that soloists employ an alternative scheme for encoding information in their spike trains; one that can be more readily read out when in addition to their spiking activity their synaptic connectivity is taken into account.

While we have demonstrated that topological featurization performs good classification for soloists, the question remains: Why does it fail for choristers? One of the defining features of choristers was their overall larger degree, leading to large associated neighborhoods. We therefore hypothesized that these larger neighborhoods were also more strongly overlapping, leading to more similar *EC* time series. As such, the information in the time series would be more redundant—each neighborhood conveying the same information. To test this, we further investigated to what extent the 25 champion samples belonging to the same parameter were overlapping and how similar their *EC* time series were. To that end, we computed the mean relative overlap of pairs of neighborhoods using the Szymkiewicz-Simpson coefficient [28], i.e. the relative number of common neurons shared by a pair of neighborhoods. Likewise, we computed the mean correlation coefficients (Pearson) between all pairs of pooled *EC* time series averaged over trials, referred to as *feature correlation*. We found a strong positive correlation between feature correlation and samples overlap in champion samples (Fig 7A), indicating that a higher number of shared neurons is linked to a higher correlation in the corresponding feature time series. Two examples of pairwise feature correlation matrices of
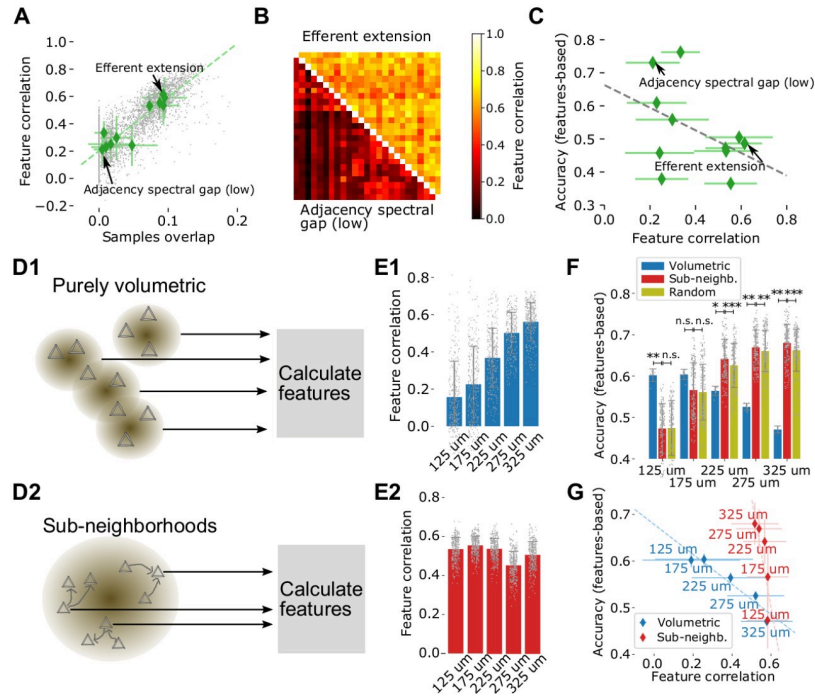
Fig 7. Feature correlation and topological featurization in volumetric samples. A: Mean relative overlap of pairs of neighborhoods that are champions of the same parameter against the mean correlation coefficient of their Euler characteristic time series (*feature correlation*). B: Pairwise feature correlations of the champions of two exemplary parameters. C: Mean feature correlation of champions against their performance in feature-based classification. D1: Feature-based classification for volumetric samples can be performed using the active sub-networks of each sample in each time step, then pooling over samples. D2: Alternatively, the 25 largest neighborhoods can be found within a single volumetric sample, which are then analyzed as other neighborhood samples. E1: Mean feature correlation for volumetric samples of various radii when the technique in D1 is employed. E2: Same, when the method outlined in D2 is employed. Only data from one volumetric sample are shown. F: Feature-based classification accuracies for volumetric samples using the purely volumetric (blue) or sub-neighborhood-based (red) techniques, or a random control for sub-neighborhoods (yellow). For sub-neighborhood-based and random conditions, we pooled data from all 25 volumetric samples. Black stars indicate Bonferroni-corrected significance (volumetric vs. sub-neighborhoods: Wilcoxon rank-sum test; sub-neighborhoods vs. random: Wilcoxon signed-rank test; n.s.: not significant, $*$: $p < 0.01$, $**$: $p < 0.001$, $***$: $p < 0.0001$). G: Feature correlation against classification accuracy for volumetric samples using the purely volumetric (blue; Pearson's $r = -0.954$, $p = 0.012$) or sub-neighborhood-based (red; Pearson's $r = -0.811$, $p = 0.096$) technique.

the parameters with the lowest and highest mean correlations respectively are illustrated (Fig 7B). Our next question was whether such a redundancy in the feature time series would lead to reduced classification accuracies. And indeed we found a negative correlation between featurization-based mean accuracies (over 6-times cross validation) and mean feature correlations (Fig 7C).

To summarize, our results align with the earlier distinction between soloists and choristers and their respective connectivity trends, adding the prediction that these two populations appear to encode information using different strategies. The soloists' strategy fails in choristers due to excessive redundancy of information. This indicates that the two codes differ in the way they employ redundancy to make the code reliable (see Discussion).

Employing the alternative decoding in volumetric samples. As before, our next aim was to apply this technique to the volumetric samples. A straight-forward approach to this simply uses the active sub-networks of 25 samples, pooling their *EC* time series (Fig 7D1). However,

as we found earlier that taking the neighborhood structure of a sample into account led to better results, we also used an alternative approach: Using only the neurons in a single volumetric sample, we extracted the 25 largest neighborhoods that could be found within a single volumetric sample, referred to as *sub-neighborhoods*, which were then analyzed as the other neighborhood samples (Fig 7D2).

In the first approach, when applying the featurization directly on volumetric samples, we found a clear dependence of the mean feature correlation on the sampling radius, with larger radii resulting in higher feature correlations (Fig 7E1). This was expected, as larger radii naturally resulted in larger overlaps of samples volumes. In the second approach, when applying featurization on sub-neighborhoods, the resulting feature correlation seemed to be rather independent of the sampling radius (Fig 7E2).

Astonishingly, when comparing the featurization-based classification accuracies of both approaches, we found that the sub-neighborhoods-based approach clearly outperformed the volumetric approach for large radii (Fig 7F). We found opposite dependencies of the mean accuracies on the sampling radii, reaching highest values for large radii in the sub-neighborhoods-based approach and for small radii in the volumetric approach respectively. So, although the sub-neighborhoods-based approach only used information extracted from a single volumetric sample instead of pooling several of them, it clearly performed better for $225\mu m$ and above. This finding cannot be fully explained by applying the featurization on smaller subsamples disregarding any neighborhood structure, as we found significantly lower performance when generating random subsamples with exactly the same sizes as the corresponding sub-neighborhoods. As before, we again found a negative correlation between accuracies and feature correlation values for both approaches (Fig 7G). Taken together, our results show that featurization can be used to decode information also in volumetric samples, with highest accuracy when taking the neighborhood structure within a volume into account.

Discussion

We have confirmed that the emergent activity of the NMC-model can be described according to the manifold hypothesis, that is, as determined by a lower-dimensional space; and that the values of those dimensions encode information about inputs given to the model. We have further demonstrated that within the morphological, electrical and synaptic diversity captured by the model, some neuronal populations are more useful in reconstructing the lower-dimensional space and using it to decode information about a stimulus. We found that these differences can be largely explained by the synaptic connectivity of neurons in the populations.

A large component of this finding was that neurons that are strongly connected to the rest of the population tend to be of the "chorister" type [27] and provide better capability for decoding stimulus identity, when using manifold-based techniques. This is not surprising, as a chorister is defined by spiking together with many other neurons and is consequently aligned with the stronger components of the underlying manifold. Thus, their spike trains carry information about those strong components.

However, we also found topological features of connectivity influencing the manifold that go beyond merely adding more synaptic connections and instead capture the specific structure of them. We demonstrated that topological parameters such as density coefficients and clustering coefficients can be used to predict the success of a stimulus classification task. The clustering coefficient measures the tendency of the connectivity to form tightly bound assemblies and the density coefficients measure the tendency to form specific types of clusters called directed simplices [29]. Their effect can be explained as follows: While synaptic input strongly constrains neuron activity, this effect is weakened by the presence of noise, especially synaptic

noise. At the same time, it has been demonstrated that correlations in synaptic input diminish the effect of noise [30]. This means that a neuron participating in connection motifs that generate correlated input, will be less subject to noise, and constrained more tightly to the manifold, improving the value of its spike train in decoding the manifold. Directed simplices have been shown to generate such correlations before [3]. The way in which other topological parameters influence classification accuracy is less clear. This is a challenging problem due to the abstract nature of the parameters. While we are not able to link the parameters to existing neuroscientific concepts at this point, this warrants follow-up study.

We found that the distinction between "well-classifying" and "poorly-classifying" neurons parallels the split into "soloists" and "choristers" that has been found before [27]. When we further investigated the emergent activity, we employed an alternative, topology-based method to reduce the dimensionality of spike trains of large neuronal populations [24]. With this "topological featurization" technique, previously poorly-classifying neurons provided sufficient information to classify stimuli with high accuracy. Topological featurization required the pooling of information from multiple neuron samples, unlike the manifold-based approach that kept them separate. Therefore, it arguably used more information for classification, potentially explaining its superior performance under some circumstances. This is however contradicted by the fact that the samples performing well with featurization were the ones that were individually small, and the larger samples performed more poorly. Further, for volumetric samples, the approach that took only a single sample instead of pooling over 25 samples had superior performance.

A code needs to provide both reliability and bandwidth. These aims stand largely in oppositions, as for a given level of noise, reliability can be increased by adding redundancy in the form of correlation, thereby reducing bandwidth. While the manifold-based technique required local connectivity that increased correlations of synaptic inputs, featurization was negatively impacted by excessive correlations. This explains its prevalence in samples of loosely connected "soloists"-type neurons. Taken together, this indicates that a qualitatively different coding scheme in soloists represents an alternative solution to the dilemma of finding a balance between the two aims. While the featurization technique cannot currently be employed *in-vivo*, as it requires information about the local connectome, advances in connectome reconstructions from electron-microscopy will allow us to validate our predictions in the future.

Does this mean that there exists neural circuitry that reads out the Euler characteristic of another population? While such a claim would certainly be premature, we at least want to consider whether such circuitry could be implemented with known biological mechanisms. The Euler characteristic is the sum of the counts of directed n-cliques for increasing n with alternating signs (see Methods). This would require, for each n-clique σ one additional readout neuron b_σ receiving input from all neurons in σ. Where b_σ should fire only when receiving input from all neurons in σ simultaneously, forming conceptually an AND conjunction and indicating that the entire n-clique is active. This could be achieved through clustering of inputs from σ onto adjacent locations on the dendrite of b_σ such that their concurrent activation triggers dendritic nonlinearities while individual activations remain local. Both cooperativity between nearby synapses [31, 32] and functional clustering of synapses with similar receptive properties [33] has been observed experimentally. Topologically, b_σ would extend the n-clique to an n + 1-clique and indeed we have found before that n-cliques are part of an unexpectedly high number of n + 1-cliques [29].

The final readout *EC* would be a neuron receiving input from all b_σs, with a sign that depends on the size (value of n) of the clique that it reads from (see S8 Fig in S1 File). That would mean that readouts from even-sized cliques would all need to be inhibitory, readouts

from uneven-sized cliques excitatory. There is no evidence for such strict motifs, nor for mechanisms that would lead to their formation. But it is possible that a less strictly alternating sum could provide similar information as the Euler characteristic. In summary, a neural readout circuitry implementing the Euler characteristic is possible, but only some of its possible building blocks have been found experimentally.

These results were obtained in a highly detailed model, albeit one of only an individual microcircuit acting alone. The nature of biological neural manifolds is likely determined as well by long-range connections and dynamic interactions between different brain regions, meaning a purely local view is limited [34, 35]. Yet, our results provide predictions on the general principles that may be employed to encode information in lower-dimensional neural manifolds.

Methods

Network simulations

Simulation methods are based on a previously published model of a neocortical microcircuit of the somatosensory cortex of the two week-old rat, here called the *NMC-model* [7]. Synaptic connectivity (with apposition-based connectome) between 219,422 neurons belonging to 55 different morphological types (m-types) was derived algorithmically starting from the appositions of dendrites and axons, and then taking into account further biological constraints such as number of synapses per connection and bouton densities [8]. Neuronal activity in the NMC-model was then simulated in the NEURON simulation environment (www.neuron.yale.edu/neuron/). Detailed information about the circuit, NEURON models and the seven connectomes of the different statistical instantiations of the NMC-model analyzed in this study are available at bbp.epfl.ch/nmc-portal/ [13].

We simulated evoked activity of the model described above for 900 seconds. After one simulated second (at $t = 0$ ms, as we discard the first second) we started applying thalamic stimuli through synapses of 2170 VPM fibers that innervated the microcircuit. The stimuli were a stream of eight different input patterns, where each pattern activated 10 randomly picked bundles (clusters) of nearby VPM fibers for 75 ms with an adapting, stochastic spiking process starting at 75 Hz and decaying to zero with a time constant of 100 ms. Bundles of VPM fibers were obtained by grouping all fibers into 100 spatially adjacent partitions using k-means clustering. The stream consisted of repetitions of the eight patterns in random order, presenting one pattern every 200 ms. Repetitions of the same stimulus used the same fibers, but with different stochastic instantiations of the spiking process. Additionally, each of the 2170 fibers was activated with a poissonian background firing rate of 0.5 Hz. As for each stimulus the pattern was chosen randomly, the total numbers of presentations differed slightly between patterns: Pattern 0: 558 repetitions; pattern 1: 559; pattern 2: 566; pattern 3: 568; pattern 4: 561; pattern 5: 560; pattern 6: 557; pattern 7; 566.

The simulation campaign was performed on 3072 processor cores of an HPE SGI 8600 supercomputer (BlueBrain5). Execution of the campaign took a total of 155 hours on this machine in 31 separate chunks. At the end of a chunk the entire internal state of the simulation, including associated random number streams, was saved and restored at the beginning of the next chunk.

Topological methods

Abstractly, the brain can be viewed as a *directed graph* $\mathcal{G} = (V, E)$, consisting of a set of *vertices* V, corresponding to the neurons, and a set of *directed edges* E. The elements of the edge set E are all pairs (i, j), where i and j are vertices in V, such that there is a synaptic connection from

neuron indexed i to neuron indexed j. Reciprocal edges between any pair of vertices i and j (namely both pairs (i, j) and (j, i)), are allowed, but no more than one edge in the same direction. Loops, namely pairs (i, i) corresponding to edges whose start and end point is the same are also excluded.

If $\mathcal{G} = (V, E)$ is a directed graph and $v \in V$ is a vertex, then the *closed neighborhood of v in* $\mathcal{G}$ is the set of all vertices in $\mathcal{G}$ that are connected to v (in either direction) and the vertex v itself. In the mathematical literature the phrase 'the closed neighborhood of v in $\mathcal{G}$' sometimes refers to the *subgraph* of $\mathcal{G}$ that is induced by the closed neighborhood of v, namely where not only the vertices v and all its neighbors, but also all the edges in $\mathcal{G}$ between pairs of vertices in the neighborhood are included. In cases where this may cause confusion, we shall attempt to clarify what we mean by referring to the closed neighborhood of v (vertices only) as the *neighborhood of v* and to the subgraph it induces as the *neighborhood graph of v*, omitting the word *closed* in both cases.

A collection of n-vertices $v_1, \ldots, v_n \in V$ of a graph in which every pair of vertices is connected by an edge is called an *n-clique*. If $\mathcal{G}$ is a directed graph then a subgraph on n vertices such that every pair of them is connected by an edge in $\mathcal{G}$, and which does not contain as a subgraph a 3-clique that is oriented cyclically, is called a *directed n-clique*. If $\mathcal{G}$ is a directed graph then one can construct a topological space that is made out of the directed n-cliques in $\mathcal{G}$. This topological space is called the *directed flag complex of* $\mathcal{G}$ [29].

If v is a vertex in a directed graph $\mathcal{G}$, then the *neighborhood graph* of v in $\mathcal{G}$ and any graph invariants one may associate with that subgraph provide graph theoretic and combinatorial information about the vertex v and its neighbors. Similarly, the directed flag complex of the neighborhood graph and associated topological invariants give higher dimensional information about v and its neighbors. Notice that since the directed flag complex of a directed graph is completely determined by the graph itself, it makes sense to talk about *topological invariants of a neighborhood* without an explicit mention of the directed flag complex, although the actual topological invariants do require us to consider the directed flag complex, as without it they cannot be defined. For a more comprehensive mathematical explanation of these concepts and the associated invariants the reader is referred to [24].

We now describe the parameters from Table 1 in a topological setting. In the S1 File we give more heuristic explanations to complement the mathematical definitions.

In-degree and out-degree. If $\mathcal{G}$ is a directed graph and v is a vertex in $\mathcal{G}$, the *in-degree of v* in $\mathcal{G}$ is the number of directed edges in $\mathcal{G}$ that end at v, and the *out-degree of v* in $\mathcal{G}$ is the number of directed edges that emanate from v in $\mathcal{G}$. The *degree of v* in $\mathcal{G}$ is the sum of its in-degree and out-degree. Note that the degree of v in $\mathcal{G}$ is one less than the number of vertices in the closed neighborhood of v. Degrees of all the vertices in a graph, and associated degree sequences, characterize properties of the graph [36]. In- and out-degree of a single vertex in a graph are, however, very local parameters, and in themselves do not inform on global structure of the graph. Note that a directed n-clique is characterized by having one vertex with in-degree zero (within the clique), called the source, and one vertex with out-degree zero (within the clique), called the sink.

Transitive clustering coefficient. The *clustering coefficient* at a vertex v in a graph $\mathcal{G}$ is a measure of how interconnected the vertices of the neighborhood of v are. This coefficient has been widely studied, initially by Watts and Strogatz [37] in undirected graphs and by Fagiolo [38] for directed graphs. We use a variation called the *transitive clustering coefficient*, which appeared in [24], and is more suitable to our context, since in our topological analysis we only consider directed cliques in the construction of the directed flag complex, and therefore in neighborhoods. The transitive clustering coefficient of a vertex v in a directed

graph $\mathcal{G}$ is defined by

$$C_2(v) = \frac{\text{the number of directed 3-cliques that contain } v}{\text{the number of possible directed 3-cliques that could contain } v}. \tag{1}$$

Any pair of edges that are incident to v could theoretically form one or two directed 3-cliques depending on their orientation. Therefore, the numerator in the definition is the number of actual directed 3-cliques that contain v, while the denominator is the maximum number of theoretically possible directed 3-cliques that contain v. The difference between this definition and that of Fagiolo is that in his definition all possible 3-cliques are considered (including the cyclical ones), where in our definition the cyclical cliques are omitted from the count. The number of possible directed 3-cliques at a vertex v in a directed graph is calculated as a function of the in-degree and out-degree of v, as well as the number of reciprocal connections it forms [24].

Topologically a 3-clique is 2-dimensional, hence the 2 in the subscript of C_2. The transitive clustering coefficient can be generalised for higher dimensions, but this is not used here.

Density coefficient. Every directed $(k + 1)$-clique contains $k + 1$ directed k-cliques as its *faces*, but no number of directed k-cliques will necessarily form any $(k + 1)$-cliques. If we let $Q_k(v)$ denote the number of directed k-cliques that contain a vertex v, then the quotient $Q_{k+1}(v)/Q_k(v)$ gives an indication of how efficiently k-cliques are put together around v to create $(k + 1)$-cliques. This number can be shown to be arbitrarily small or large, depending on the ambient graph [24]. Therefore, to obtain a distribution as a number between 0 and 1, we normalise and define the k-th density coefficient for a vertex v *in a graph $\mathcal{G}$ with n vertices* to be

$$D_k(v) = \frac{k}{(k+1)(n-k)} \cdot \frac{Q_{k+1}(v)}{Q_k(v)}. \tag{2}$$

With this definition, if $\mathcal{G}$ is a directed graph on n vertices, where any pair of vertices is reciprocally connected then the k-th density coefficient of any vertex in it will be 1 for all k. For instance the 2nd density coefficient of a vertex v is defined by

$$D_2(v) = \frac{2}{3(n-2)} \cdot \frac{Q_3(v)}{\deg(v)}.$$

In all cases the subscript k refers to $(k + 1)$-cliques being topologically k-dimensional. Notice also that $Q_2(v) = \deg(v)$.

Homological parameters. If $\mathcal{G}$ is a directed graph, let $X_{\mathcal{G}}$ denote its directed flag complex. This is a topological space obtained from $\mathcal{G}$ by replacing every directed 2-clique (a directed edge) by a line segment (or in mathematical language—a 1-simplex), every directed 3-clique by a solid triangle (a 2-simplex), every directed 4-clique by a solid tetrahedron (a 3-simplex), and so on in higher dimension. In general each directed $(n + 1)$-clique is replaced by an n-simplex—an n-dimensional object that generalises the concepts of, line segment, triangle and tetrahedron to a general dimension. The *directed flag complex* $X_{\mathcal{G}}$ of the directed graph $\mathcal{G}$ has been defined and studied in the context of neuronal networks and neuroscience in [29, 39].

Any directed $(n + 1)$-clique contains exactly $n + 1$ directed n-sub-cliques. The corresponding n-simplex contains $n + 1$ $(n - 1)$-*dimensional faces*, that are themselves $(n - 1)$-simplices in $X_{\mathcal{G}}$. Using this structure it is possible to associate with the directed flag complex an algebraic object called a *chain complex*. It is constructed as follows. For every $n \geq 0$, take $C_n(\mathcal{G})$ to be the vector space over the finite field of 2 elements (or any other field) [40], with basis given abstractly by the set of all n-simplices in $X_{\mathcal{G}}$ or equivalently all the directed $(n + 1)$-cliques in $\mathcal{G}$.

Taking faces of simplices gives rise, for every $n \geq 1$, to a linear transformation called a differential

$$\partial_n : C_n(\mathcal{G}) \to C_{n-1}(\mathcal{G}).$$

The *n-the homology group* of $X_\mathcal{G}$ (with coefficients in the field of 2 elements) is defined by

$$H_n(X_\mathcal{G}) = \mathrm{Ker}(\partial_n)/\mathrm{Im}(\partial_{n+1}),$$

namely the quotient vector space of the kernel of ∂_N by the subspace that is the image of ∂_{n+1}. For the purpose of the discussion in this article, what is important to understand is only the dimension of the vector space $H_n(X_\mathcal{G})$—the so called the *n*-th Betti number of $X_\mathcal{G}$. If we let D_n denote the matrix representing the linear transformation ∂_n, then the *n*-th Betti number, denoted b_n or $b_n(\mathcal{G})$, is calculated as the difference

$$b_n = \mathrm{Null}(D_n) - \mathrm{rk}(D_{n+1}),$$

i.e., the dimension of the null-space D_n minus the dimension of the column space of D_{n+1}. These numbers depend on the field as well as on the graph $\mathcal{G}$, but in this article we worked only with the field of 2 elements. The interested reader is referred to [41] for a thorough mathematical background on these concepts.

The 0-th Betti number b_0 is exactly the number of connected components of $\mathcal{G}$. The 1-st Betti number b_1 can be thought of as the number of distinct loops in $\mathcal{G}$ that are not bounding a 2-dimensional subspace in $X_\mathcal{G}$. Intuitively, the Betti numbers of $X_\mathcal{G}$ are a count of *n*-dimensional "cavities" in $X_\mathcal{G}$ [29].

In this paper we consider two extra topological metrics that are associated to Betti numbers. The first is the classical *Euler characteristic*. The Euler characteristic $\chi(X_\mathcal{G})$ can be computed in two ways that yield the same number. One is the alternating sum

$$\chi = b_0 - b_1 + b_2 - b_3 + \cdots$$

of the Betti numbers of $X_\mathcal{G}$. The other is the alternating sum of the number of directed cliques in each dimension in $\mathcal{G}$. The Euler characteristic is relatively very easy to compute (using the second method) and although it is considered to be a weak invariant of topological spaces, it is frequently extremely efficient both in theory and in applications. It is used in a classification task in [24] and as a topological parameter in this article.

The second topological parameter we use here is the *normalized Betti coefficient*. It is a weighted sum of the Betti numbers:

$$\mathfrak{B}(X_\mathcal{G}) := \sum_{i=0}^{\infty} \frac{(i+1) \cdot b_i}{\text{the number of } i\text{-dimensional simplices in } X_\mathcal{G}}. \tag{3}$$

The Betti coefficient is a rough measure of how efficient the graph $\mathcal{G}$ is in creating cavities in all dimensions where they exist in the directed flag complex $X_\mathcal{G}$.

Spectral parameters. Every square $n \times n$ real valued matrix A has *eigenvalues* $\{\lambda_1, \ldots, \lambda_n\}$, which are the (real or complex) solutions to the *characteristic equation* $A\mathbf{x} = \lambda\mathbf{x}$ where $\mathbf{x}$ is a vector of length n. Equivalently the eigenvalues are the roots of the *characteristic polynomial of A*. The collection of eigenvalues of a matrix A is often referred to as the *spectrum* of A.

Considering the set of moduli (absolute values) of the eigenvalues of A, one associates three invariants with A:

- The *spectral radius* of A is the largest modulus of an eigenvalue.

- The *low spectral gap* of A is the smallest modulus of a non-zero eigenvalue of A.

- The *high spectral gap* of A is the difference between the moduli of its two largest eigenvalues (sorted by their moduli).

Let $\mathcal{G}$ be a directed graph with n vertices. The *adjacency matrix* A of $\mathcal{G}$ has (i,j)-entry $a_{i,j} = 1$ if there is a directed edge from vertex i to vertex j, and 0 otherwise. The *Chung–Laplacian matrix* has a more involved definition, which can be found in [42]. A graph is said to be *strongly connected* if for any two distinct vertices u and v in $\mathcal{G}$ there is a directed path in $\mathcal{G}$ from u to v. The Chung–Laplacian matrix is only defined on a strongly connected graph. In this paper we consider the spectral radius of the adjacency matrix of the graphs we studied, as well as its low and high spectral gaps. We also considered the Chung–Laplacian spectral gaps of the largest strongly connected components of graphs, whenever these can be uniquely determined.

These invariants are well studied in theoretical and applied graph theory, where their main objective is to relate various structural properties of graphs to the spectrum. For example the Laplacian spectral gap is famously known to measure how easy the graph is to disconnect through so-called Cheeger inequalities [43].

Relative boundary.   For a subset $S \subseteq V$ of vertices in a graph $\mathcal{G} = (V, E)$, the *edge boundary*

$$\partial S := \{(i,j) \in E \mid \text{ either } i \in S, j \notin S \text{ or } j \in S, i \notin S\} \tag{4}$$

of S consists of all edges with exactly one endpoint in S. Likewise, the *edge volume*

$$|S| := |\{(i,j) \in E \mid i \in S, j \in S\}| \tag{5}$$

of S is the number of edges whose both endpoints are in S. We then define the *relative boundary*

$$rb(S) := \frac{|\partial S|}{|S|} \tag{6}$$

of S as the size of its edge boundary divided by its volume. Relative boundary is related to the so-called isoperimetric or Cheeger number [43] and is designed to measure how strongly the subset S connects to the rest of the graph relative to its internal connectivity.

Extension.   For a subset $S \subseteq V$ of vertices in a graph $\mathcal{G} = (V, E)$, its extension is the number of vertices that are connected to S but are not in S itself. The *afferent extension* of S and the *efferent extension* of S are defined as

$$\text{ae}(S) := |\{v \in V \backslash S \mid (v, j) \in E, \ j \in S\}|, \tag{7}$$

$$\text{ee}(S) := |\{v \in V \backslash S \mid (i, v) \in E, \ i \in S\}|, \tag{8}$$

respectively. Note that, for example, a vertex v outside of S for which $(v, s_1), (v, s_2) \in E$ for distinct $s_1, s_2 \in S$ will be counted only once in ae(S), so as to distinguish the extension of S from the edge boundary of S. In other words, ae(S) + ee(S) is bounded above by the edge boundary of S, and the bound is attained whenever no vertex in $V \setminus S$ is connected by more than one edge to S. In that case, the edge boundary coincides with the extension.

Structural analysis of the model

The results of the simulations were analyzed according to the pipeline depicted in Fig 8. Inputs were:

1. The spike trains of the 31,346 most central excitatory neurons in the simulations

2. For each neuron its morphological type, layer and location in the model (x,y,z-coordinates)
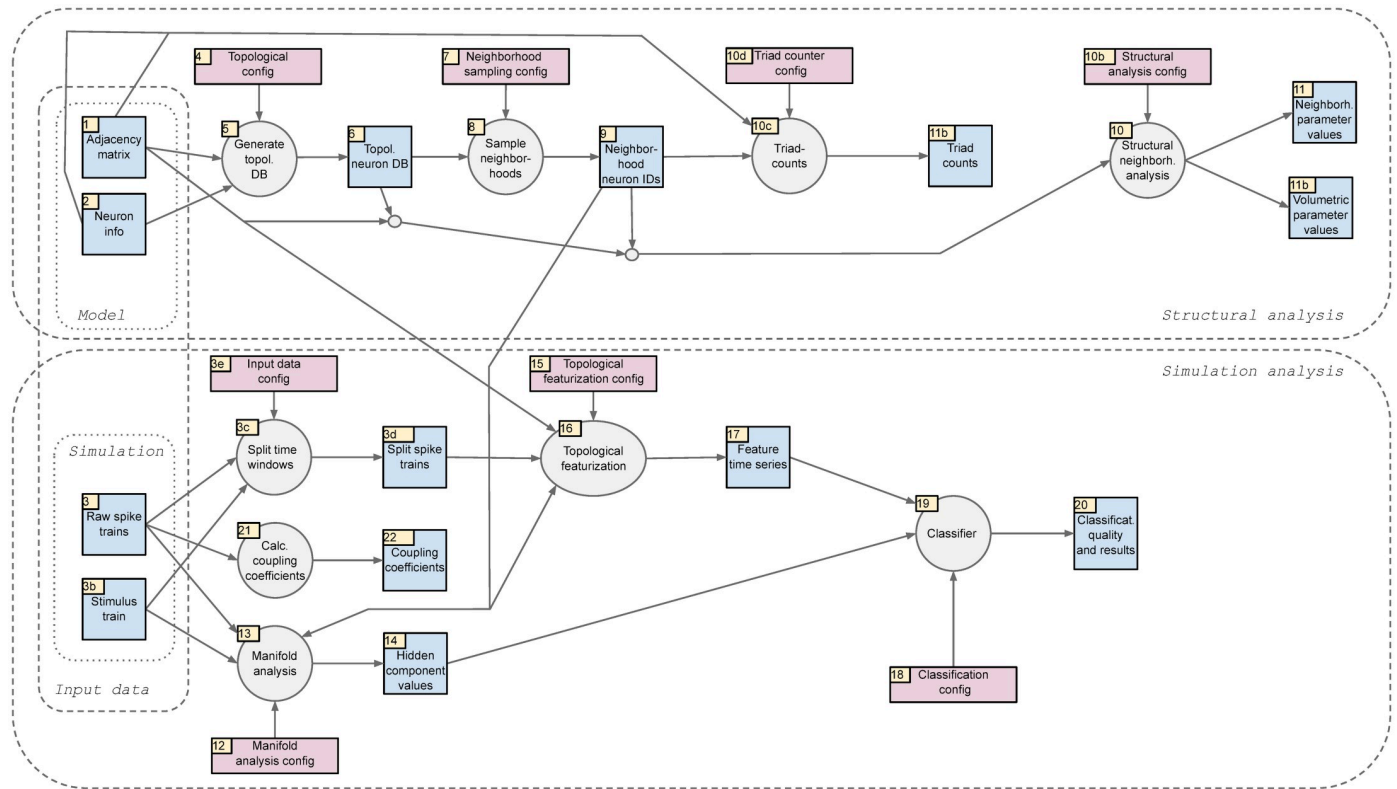
Fig 8. Overview of the inputs into the analysis pipeline and its individual analysis steps.

3. The adjacency matrix of synaptic connections between all neurons in the model

4. The identifier of the pattern presented during each stimulation (the "stimulus stream").

These inputs can be downloaded from https://doi.org/10.5281/zenodo.4317336. The analysis pipeline conceptualized in Fig 8 was implemented in custom python code (python version 3.7), except for parts of the topological analysis which used custom c++ code that was wrapped for use in python using pybind11. The entirety of the pipeline can be obtained from: https://github.com/BlueBrain/topological_sampling.

The following sections will explain the individual steps of the pipeline, beginning with the structural analysis of the model and its connectivity, followed by the analysis of the simulated neuron activity.

Generating neuron samples. Neuron samples were generated in one of six ways:

1. Volumetric. First, we found the center of the neuron population by averaging their x, y, and z-coordinates. Next, we added a random offset between $-100\mu m$ and $100\mu m$ for the x- and z-coordinates (parallel to layer boundaries) and $-300\mu m$ and $300\mu m$ for the y-coordinate (orthogonal to layer boundaries). Next, we found all neurons in the model within a certain radius of that point and randomly picked 600 neurons from them. We repeated these steps to generate 25 samples each for radii of $125\mu m$, $175\mu m$, $225\mu m$, $275\mu m$ and $325\mu m$.

2. Champions. We generated champions by finding the 25 neighborhoods that yielded the largest values for one of the topological parameters described above. We limited the selection to neighborhoods with at least 50 neurons.

3. Random samples. For each of the morphological types of neurons in the model, we randomly picked 25 neurons and used their associated neighborhoods.

4. Subsampled. For the champions of in-degree, out-degree, and Euler characteristic we generated random subsamples. First, we randomly picked five champions of each of the three parameters. Next, we randomly selected a certain fraction of the neurons contained in the neighborhoods. We repeated this five times, generating five subsamples of each picked neighborhood. We thus generated subsamples at 90%, 70%, 50%, 25% and 15% of the original neighborhood sizes.

5. Sub-neighborhoods. For each volumetric sample, we picked the 25 largest neighborhoods contained within a sample.

6. Random sub-neighborhoods. As a control condition, we generated random subsamples from each volumetric sample which have exactly the same sizes as their corresponding 25 largest sub-neighborhoods.

Triad counts. Input into the triad counter was the adjacency matrix of the model and the identifiers of neurons in a given sample. First we extracted the submatrix defining the internal connectivity of the sample. Then we recursively iterated through all possible combinations of three neurons of the sample and classified their connectivity into one of 13 motifs. We limited this analysis to neuron triplets that were strongly connected in either direction, otherwise three additional motifs (unconnected, single connection, single bidirectional connection) would have been possible.

We then calculated the expected numbers of each motif in two control models. First, an Erdos-Renyi graph with the same number of nodes and edges, yielding $C_{ER}(i)$, the expected count of motif i. Second, an Erdos-Renyi graph with the same number of nodes and edges, but taking into account the sampling procedure used. That is, taking into account that the sample contains one central neuron and its neighborhood. We calculated this as follows.

First, we calculated the number of triplets that contain the center (central neuron):

$$T_{center} = \frac{(N-1) \cdot (N-2)}{2},\qquad(9)$$

Where N refers to the size of the sample. In this arrangement, the connections binding the center to the two other neurons can have one of three patterns. From the view of the center they can be efferent-efferent (probability 0.25), efferent-afferent (probability 0.5) or afferent-afferent (0.25). For each of these possibilities we analytically derived the expected numbers of motifs assuming that the remaining connections were subject to a uniform, statistically independent probability derived. This included the option to turn the connection binding the center to a neighborhood member into a bidirectional connection. The thusly derived motif probabilities, P_{center}, multiplied by the number of center-including triplets yielded the first part of the expected motif counts.

Next, the number of triplets not containing the center was:

$$T_{non-center} = \frac{(N-1) \cdot (N-2) \cdot (N-3)}{6}\qquad(10)$$

For these we derived the expected motif counts according to an Erdos-Renyi control. Total expected count of motif i in neighborhood sampling was then:

$$C_{neighborhood}(i) = T_{center}(i) * P_{center}(i) + T_{non-center}(i) * P_{ER}(i)\qquad(11)$$

The degree of over- and under-expression of motif i in volumetric samples was then calculated as:

$$O_{vol}(i) = \frac{C_{vol}(i) - C_{ER}(i)}{C_{vol}(i) + C_{ER}(i)}. \tag{12}$$

The degree of over- and under-expression in champion samples was calculated relative to the volumetric samples. First we normalized each sample against their respective control:

$$\hat{O}_{vol}(i) = \frac{C_{vol}(i)}{C_{ER}(i)}; \hat{O}_{champ}(i) = \frac{C_{champ}(i)}{C_{neighborhood}(i)}, \tag{13}$$

Then, we normalized the motif count in the samples to the mean and standard deviation of the volumetric samples:

$$O_{champ}(i) = \frac{\hat{O}_{champ}(i) - mean(\hat{O}_{vol}(i))}{std(\hat{O}_{vol}(i))}, \tag{14}$$

Where *mean* and *std* refer to the mean and standard deviation over volumetric samples.

Calculating topological parameters for samples. We calculated the values of the topological parameters for all possible neighborhoods, i.e. we considered each neuron in the model as a center and calculated the parameters of the resulting neighborhood. We also calculated parameter values for volumetric samples in two ways.

First, by simply applying the topological method to the connectivity of the sample. This could be done for all parameters except in-degree, out-degree, transitive clustering coefficient. These parameters were calculated in neighborhood samples as the in-degree, out-degree, etc. *of the center* and were consequently undefined for volumetric samples. We instead used the mean in-degree, out-degree of clustering coefficient of all neurons in the volumetric sample.

Second, by taking the neighborhood structure of the sample into account. To that end, we first calculated the value of a given parameter for all possible neighborhoods in the model as described above. Next, we calculated for each neighborhood the relative size of its overlap with the volumetric sample in question, i.e. the number of neurons contained in both divided by the number of neurons contained in the sample. We then selected the n neighborhoods with the largest overlap and calculated the weighted mean of their value of the parameter, with weights proportional to the overlap. Note that the center of a neighborhood did not have to be part of the sample to let it impact the weighted mean. We optimized the value of n to yield the best predictor of accuracy, separately for each topological parameter.

Analysis of simulated neuron activity

Calculation of coupling coefficients. Input into this step were the raw spike trains of all neurons recorded in the simulations.

We started by binning the spike trains of all neurons in the model into 10 ms bins. This yielded a $N \times T$ sparse matrix, where N was the number of neurons and T the number of time bins and the entry at i, j specified the number of spikes of neuron i in time bin j. For the coupling coefficient of neuron i, we used the ith row of the matrix as the time series of firing of that neuron, and the mean value across all rows of the matrix, except i as the average firing rate of all others. We then calculated the coupling coefficient as the normalized correlation between the two time series (numpy.corrcoef).

Dimensionality reduction through factor analysis. Input into this step were the raw spike trains of all neurons recorded in the simulations, information about the pattern identity of each stimulation, and the identifiers of a sample of neurons.

We began by binning the spike trains of all samples neurons into 10 ms bins. This yielded a $N_{sample} \times T$ sparse matrix, where N_{sample} denotes the number of neurons in the sample. Non-spiking neurons were removed. Next, we extracted the twelve strongest components from this N_{sample}-dimensional time series using factor analysis (sklearn.decomposition.FactorAnalysis). Then we split the resulting 12-dimensional time series into 200 ms time windows that each corresponded to a stimulus presentation. Finally, we grouped these time windows by the identity of the stimulus pattern presented at that time.

Topological featurization. For a given type of neuron sample, based on a topological parameter or randomly sampled (see "generating neuron samples" above), we considered all 25 generated neighborhoods. Spike trains for each of the individual neurons in the microcircuit and each stimulus were binned into 10 ms time bins. In each time bin we recorded the neurons that fired at least one spike, i.e. active neurons. For each of the selected neighborhoods, we now computed the directed flag complex of the subgraph of the neighborhood graph induced by the active neurons in each time bin, and then we computed the Euler characteristic (EC) of that directed flag complex. For each stimulus, this reduced the activity of the neurons in a neighborhood to a time series of 20 (one per time bin) EC values. The series for all 25 neighborhoods were concatenated and used used as feature vectors for our classification. In [24] the same method was used, but many other topological and graph invariants were used there as feature parameters, and not only Euler characteristic. However, Euler characteristic was shown there to be one of the strongest performing feature parameters in terms of classification accuracy.

Stimulus classification. Input into the classifier was either: The twelve strongest components of the spiking activity of a neighborhood, extracted as detailed above. Or: The time series of Euler characteristic values of the active sub-neighborhood extracted from the activity of 25 neighborhoods as detailed in the main text. In both cases, the input was split into time windows that each represented a single trial, i.e. a single stimulus. The time windows were further grouped by the identity of the stimulus pattern used in the trial.

Next, we generated for each time window a time series of expected outputs of the classifier. This was simply the identity of the stimulus (an integer between 0 and 7), repeated for each time step. 60% of the trials and their associated expected outputs were used to train a linear classifier (sklearn.svm.SVC). The remaining 40% were used to determine classification accuracy. This was repeated 5 times, thereby conducting 6-times cross validation.

Dependence of classification accuracy on topological parameters. We first generated a model of the dependence of accuracy on neighborhood sizes in a sample. To that end, we iterated over each neuron in a sample, considered the neuron as the center of its own neighborhood and calculated the size of that neighborhood. We call the mean of these neighborhood sizes for a sample the *neighborhood size average (NSA)*. This measure could be calculated equally for both neighborhood and volumetric samples. Next, as the model of the impact of NSA, we conducted a linear fit of NSA against classification accuracy (using the manifold method) based on the data of 1276 randomly picked neighborhoods. The fit minimized the sum of squared errors and was performed using the statsmodels package in python 3.7.

We calculated the *residual accuracy* for each sample—neighborhood or volumetric—by subtracting the prediction of this model from the classification accuracy values. We then determined the relation between topological parameters and residual accuracy as follows. We began with a simple control model predicting the residual accuracy from the sampling radius in the case of volumetric samples or from the morphological type of the center in the case of

neighborhood samples. Both morphological type and radius were considered categorical variables. Again, the fit was performed using statsmodels in python.

Next, we normalized the values of each topological parameter to zero mean and unity variance and generated linear models of the residual accuracy taking into account the effects of both radius / morphological type as a categorical variable and the normalized value of the parameter. We created such a model for each topological parameter.

We then calculated the slope of the fits in the parameter to assess the strength of the effect of that parameter; and the fraction of variance explained, from which we subtracted the fraction of variance explained by the simpler control models.

Supporting information

S1 File. Supplementary explanations and figures.
(PDF)

S1 Fig. A: Graphs, neighborhoods, and cliques. B: Different ways to complete two edges to a directed 3-clique. C: Different types of graphs. Note the Chung–Laplacian spectrum considers only the largest strongly connected component when computing eigenvalues. D: Comparisons of four different graph parameters relative to one of its vertices. E: Comparisons of three different graph parameters and two different spectra (unique absolute values of eigenvalues of matrices).
(PDF)

S2 Fig. Comparisons of the edge boundary, relative boundary, afferent extension, and efferent extension for particular subsets. Subsets are chosen as neighborhoods.
(PDF)

S3 Fig. A: All investigated topological parameters, with pairs that are mutually redundant (in terms of resulting triad motif expression patterns) highlighted in yellow. B: For the champion neighborhoods of the non-redundant parameters (columns), we consider the values of all parameters (rows), normalized in terms of the percentile of the overall distribution of said parameter (see color bar).
(PDF)

S4 Fig. Time series of the twelve strongest components (panels) during presentation of the individual stimulus patterns (colored traces). Thick lines and error bars: mean and SEM. Thin lines: for five randomly selected trials using a given pattern.
(PDF)

S5 Fig. Classification accuracies, using the manifold-based method, for the randomly selected neighborhoods with centers of the indicated morphological type. Grey bars and error bars: mean and std. Blue dots: individual neighborhoods.
(PDF)

S6 Fig. Values of topological parameters against residual accuracy for randomly selected neighborhoods. Grey dots: individual neighborhoods. Black line: linear fit.
(PDF)

S7 Fig. A: Synthetic values of topological parameters for volumetric samples against their residual accuracy. Blue dots: individual samples. Black line: linear fit. B: Number of neighborhoods used in the calculation of the synthetic values (see Section Calculating topological parameters for samples) against the resulting correlation (pearsonr) with classifier accuracy. Individual, colored lines: For individual topological parameters. Colored dots: Maxima of the

absolute value of correlations, indicating the number of neighborhoods used in the remainder of the manuscript.
(PDF)

Author Contributions

Conceptualization: Michael W. Reimann, Henri Riihimäki, Ran Levi.

Investigation: Michael W. Reimann, Henri Riihimäki, Jason P. Smith, Christoph Pokorny.

Methodology: Michael W. Reimann, Henri Riihimäki, Jason P. Smith.

Software: Michael W. Reimann, Henri Riihimäki, Jason P. Smith, Jānis Lazovskis, Christoph Pokorny.

Supervision: Michael W. Reimann, Ran Levi.

Validation: Michael W. Reimann.

Visualization: Michael W. Reimann, Jānis Lazovskis, Christoph Pokorny.

Writing – original draft: Michael W. Reimann.

Writing – review & editing: Michael W. Reimann, Henri Riihimäki, Jānis Lazovskis, Christoph Pokorny, Ran Levi.

References

1. Gallego JA, Perich MG, Miller LE, Solla SA. Neural Manifolds for the Control of Movement. Neuron. 2017; 94(5):978–984. https://doi.org/10.1016/j.neuron.2017.05.025 PMID: 28595054

2. Tkacik G, Prentice JS, Balasubramanian V, Schneidman E. Optimal population coding by noisy spiking neurons. Proceedings of the National Academy of Sciences. 2010; 107(32):14419–14424. https://doi.org/10.1073/pnas.1004906107 PMID: 20660781

3. Nolte M, Reimann MW, King JG, Markram H, Muller EB. Cortical reliability amid noise and chaos. Nature Communications. 2019; 10(1):3792. https://doi.org/10.1038/s41467-019-11633-8 PMID: 31439838

4. Tolhurst DJ, Movshon JA, Dean AF. The statistical reliability of signals in single neurons in cat and monkey visual cortex. Vision Research. 1983; 23(8):775–785. https://doi.org/10.1016/0042-6989(83)90200-6 PMID: 6623937

5. Stern EA, Kincaid AE, Wilson CJ. Spontaneous Subthreshold Membrane Potential Fluctuations and Action Potential Variability of Rat Corticostriatal and Striatal Neurons In Vivo. Journal of Neurophysiology. 1997; 77(4):1697–1715. https://doi.org/10.1152/jn.1997.77.4.1697 PMID: 9114230

6. Shadlen MN, Newsome WT. The Variable Discharge of Cortical Neurons: Implications for Connectivity, Computation, and Information Coding. The Journal of Neuroscience. 1998; 18(10):3870–3896. https://doi.org/10.1523/JNEUROSCI.18-10-03870.1998 PMID: 9570816

7. Markram H, Muller E, Ramaswamy S, Reimann MW, Abdellah M, Sanchez CA, et al. Reconstruction and Simulation of Neocortical Microcircuitry. Cell. 2015; 163(2):456–492. https://doi.org/10.1016/j.cell.2015.09.029 PMID: 26451489

8. Reimann MW, King JG, Muller EB, Ramaswamy S, Markram H. An algorithm to predict the connectome of neural microcircuits. Frontiers in Computational Neuroscience. 2015; 9:120. https://doi.org/10.3389/fncom.2015.00120 PMID: 26500529

9. Gal E, London M, Globerson A, Ramaswamy S, Reimann MW, Muller E, et al. Rich cell-type-specific network topology in neocortical microcircuitry. Nature Neuroscience. 2017; 20(7):1004–1013. https://doi.org/10.1038/nn.4576 PMID: 28581480

10. Reyes-Puerta V, Sun JJ, Kim S, Kilb W, Luhmann HJ. Laminar and Columnar Structure of Sensory-Evoked Multineuronal Spike Sequences in Adult Rat Barrel Cortex In Vivo. Cerebral Cortex (New York, NY: 1991). 2015; 25(8):2001–2021. https://doi.org/10.1093/cercor/bhu007 PMID: 24518757

11. Renart A, de la Rocha J, Bartho P, Hollender L, Parga N, Reyes A, et al. The asynchronous state in cortical circuits. Science (New York, NY). 2010; 327(5965):587–590. https://doi.org/10.1126/science.1179850

12. Luczak A, Bartho P, Marguet SL, Buzsaki G, Harris KD. Sequential structure of neocortical spontaneous activity in vivo. Proceedings of the National Academy of Sciences. 2007; 104(1):347–352. https://doi.org/10.1073/pnas.0605643104 PMID: 17185420

13. Ramaswamy S, Courcol JD, Abdellah M, Adaszewski SR, Antille N, Arsever S, et al. The neocortical microcircuit collaboration portal: a resource for rat somatosensory cortex. Frontiers in Neural Circuits. 2015; 9:44. https://doi.org/10.3389/fncir.2015.00044 PMID: 26500503

14. Stevens CF, Wang Y. Facilitation and depression at single central synapses. Neuron. 1995; 14(4):795–802. https://doi.org/10.1016/0896-6273(95)90223-6 PMID: 7718241

15. Markram H, Tsodyks M. Redistribution of synaptic efficacy between neocortical pyramidal neurons. Nature. 1996; 382(6594):807–810. https://doi.org/10.1038/382807a0 PMID: 8752273

16. Abbott LF. Synaptic Depression and Cortical Gain Control. Science. 1997; 275(5297):221–224. https://doi.org/10.1126/science.275.5297.221 PMID: 8985017

17. Simkus CRL, Stricker C. Properties of mEPSCs recorded in layer II neurones of rat barrel cortex. The Journal of Physiology. 2002; 545(2):509–520. https://doi.org/10.1113/jphysiol.2002.022095 PMID: 12456830

18. Ling DSF, Benardo LS. Restrictions on Inhibitory Circuits Contribute to Limited Recruitment of Fast Inhibition in Rat Neocortical Pyramidal Cells. Journal of Neurophysiology. 1999; 82(4):1793–1807. https://doi.org/10.1152/jn.1999.82.4.1793 PMID: 10515969

19. Ribrault C, Sekimoto K, Triller A. From the stochasticity of molecular processes to the variability of synaptic transmission. Nature Reviews Neuroscience. 2011; 12(7):375–387. https://doi.org/10.1038/nrn3025 PMID: 21685931

20. Faisal AA, Selen LPJ, Wolpert DM. Noise in the nervous system. Nature Reviews Neuroscience. 2008; 9(4):292–303. https://doi.org/10.1038/nrn2258 PMID: 18319728

21. Meyer HS, Wimmer VC, Hemberger M, Bruno RM, D Kock P j C, et al. Cell Type–Specific Thalamic Innervation in a Column of Rat Vibrissal Cortex. Cerebral Cortex. 2010; 20(10):2287–2303. https://doi.org/10.1093/cercor/bhq069 PMID: 20534783

22. Desbordes G, Jin J, Alonso JM, Stanley GB. Modulation of Temporal Precision in Thalamic Population Responses to Natural Visual Stimuli. Frontiers in Systems Neuroscience. 2010; 1. https://doi.org/10.3389/fnsys.2010.00151 PMID: 21151356

23. Petersen RS, Brambilla M, Bale MR, Alenda A, Panzeri S, Montemurro MA, et al. Diverse and Temporally Precise Kinetic Feature Selectivity in the VPm Thalamic Nucleus. Neuron. 2008; 60(5):890–903. https://doi.org/10.1016/j.neuron.2008.09.041 PMID: 19081382

24. Conceição P, Govc D, Lazovskis J, Levi R, Riihimäki H, Smith JP. An application of neighbourhoods in digraphs to the classification of binary dynamics. arXiv:210406519 [math]. 2021.

25. Riihimäki H, Chachólski W, Theorell J, Hillert J, Ramanujam R. A topological data analysis based classification method for multiple measurements. BMC Bioinformatics. 2020; 21(1):336. https://doi.org/10.1186/s12859-020-03659-3 PMID: 32727348

26. Petilla Interneuron Nomenclature Group, Ascoli GA, Alonso-Nanclares L, Anderson SA, Barrionuevo G, Benavides-Piccione R, et al. Petilla terminology: nomenclature of features of GABAergic interneurons of the cerebral cortex. Nature Reviews Neuroscience. 2008; 9(7):557–568. https://doi.org/10.1038/nrn2402 PMID: 18568015

27. Okun M, Steinmetz NA, Cossell L, Iacaruso MF, Ko H, Barthó P, et al. Diverse coupling of neurons to populations in sensory cortex. Nature. 2015; 521(7553):511–515. https://doi.org/10.1038/nature14273 PMID: 25849776

28. Vijaymeena MK, Kavitha K. A Survey on Similarity Measures in Text Mining. Machine Learning and Applications: An International Journal. 2016; 3(1):19–28. https://doi.org/10.5121/mlaij.2016.3103

29. Reimann MW, Nolte M, Scolamiero M, Turner K, Perin R, Chindemi G, et al. Cliques of Neurons Bound into Cavities Provide a Missing Link between Structure and Function. Frontiers in Computational Neuroscience. 2017; 11:48. https://doi.org/10.3389/fncom.2017.00048 PMID: 28659782

30. Wang HP, Spencer D, Fellous JM, Sejnowski TJ. Synchrony of Thalamocortical Inputs Maximizes Cortical Reliability. Science. 2010; 328(5974):106–109. https://doi.org/10.1126/science.1183108 PMID: 20360111

31. Harnett MT, Makara JK, Spruston N, Kath WL, Magee JC. Synaptic amplification by dendritic spines enhances input cooperativity. Nature. 2012; 491(7425):599–602. https://doi.org/10.1038/nature11554 PMID: 23103868

32. Weber JP, Andrasfalvy BK, Polito M, Mago A, Ujfalussy BB, Makara JK. Location-dependent synaptic plasticity rules by dendritic spine cooperativity. Nature Communications. 2016; 7(1):11380. https://doi.org/10.1038/ncomms11380 PMID: 27098773

33. Iacaruso MF, Gasler IT, Hofer SB. Synaptic organization of visual space in primary visual cortex. Nature. 2017; 547(7664):449–452. https://doi.org/10.1038/nature23019 PMID: 28700575

34. Allen WE, Chen MZ, Pichamoorthy N, Tien RH, Pachitariu M, Luo L, et al. Thirst regulates motivated behavior through modulation of brainwide neural population dynamics. Science (New York, NY). 2019; 364(6437):253. https://doi.org/10.1126/science.aav3932 PMID: 30948440

35. Stringer C, Pachitariu M, Steinmetz N, Reddy CB, Carandini M, Harris KD. Spontaneous behaviors drive multidimensional, brainwide activity. Science. 2019; 364(6437):eaav7893. https://doi.org/10.1126/science.aav7893 PMID: 31000656

36. Diestel R. Graph Theory. 5th ed. Springer; 2017.

37. Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. Nature. 1998; 393(6684): 440–442. https://doi.org/10.1038/30918 PMID: 9623998

38. Fagiolo G. Clustering in complex directed networks. Physical Review E. 2007; 76(2):026107. https://doi.org/10.1103/PhysRevE.76.026107 PMID: 17930104

39. Masulli P, Villa AEP. The topology of the directed clique complex as a network invariant. SpringerPlus. 2016; 5(1):388. https://doi.org/10.1186/s40064-016-2022-y PMID: 27047714

40. Lidl R, Niederreiter H. Finite Fields. In: Encyclopedia of Mathematics and Its Applications. vol. 20. 2nd ed. Cambridge University Press; 1997.

41. Hatcher A. Algebraic Topology. Cambridge University Press; 2002. Available from: http://pi.math.cornell.edu/~hatcher/AT/ATpage.html.

42. Chung F. Laplacians and the Cheeger Inequality for Directed Graphs. Annals of Combinatorics. 2005; 9 (1):1–19. https://doi.org/10.1007/s00026-005-0237-z

43. Chung F. Spectral Graph Theory. vol. 92 of CBMS Regional Conference Series in Mathematics. American Mathematical Society; 1996. Available from: http://www.ams.org/cbms/092.