Using persistent homology to reveal hidden covariates in systems governed by the kinetic Ising model

Gard Spreemann,[1,*] Benjamin Dunn,[2] Magnus Bakke Botnan,[1,†] and Nils A. Baas[1]
[1]*Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim 7491, Norway*
[2]*Kavli Institute for Systems Neuroscience, Norwegian University of Science and Technology, Trondheim 7491, Norway*

(Received 2 October 2017; published 26 March 2018)

We propose a method, based on persistent homology, to uncover topological properties of *a priori* unknown covariates in a system governed by the kinetic Ising model with time-varying external fields. As its starting point the method takes observations of the system under study, a list of suspected or known covariates, and observations of those covariates. We infer away the contributions of the suspected or known covariates, after which persistent homology reveals topological information about unknown remaining covariates. Our motivating example system is the activity of neurons tuned to the covariates physical position and head direction, but the method is far more general.

DOI: 10.1103/PhysRevE.97.032313

I. INTRODUCTION

The resolution of neural recordings continues to improve rapidly, generating increased interest in the use of appropriate models to fit to the data. Identification of the relevant features to include is an important but often puzzling step in the process, in particular when the possibilities are many and difficult to observe [1,2]. In this work we focus on time-varying features, or covariates, and describe a method for revealing their topological properties, and how this can be used to guide model construction.

The modeling framework we consider is that of the generalized linear model (GLM) [3], specifically the kinetic Ising model. We assume that these covariates come from (non-time-dependent) functions on the system's state space. For example, in the case of head direction or in the case of place cells, the relevant covariates would arise from the non-time-dependent selectivity of the neuron to the state space underlying the selectivity, i.e., the head direction or position, respectively, of the animal. Recent studies have stressed the importance of correctly identifying and including relevant covariates when modeling a wide array of systems [4]. However, the number of well-characterized cell types, such as head direction or place cells, is minuscule relative to the likely number of cell types.

Potential issues from ignoring covariates include the following:

(i) incorrectly attributing selectivity that is related to but different from the activity of the cell, i.e., head direction versus movement direction of the animal;

(ii) failure to identify additional unknown tuning of some or all of the recorded neurons;

(iii) incorrect identification of functional connectivity that is more likely attributed to common selectivity of the neurons.

The first and third issues mentioned here can be approached using common model selection approaches [5,6]. The second issue, however, is less straightforward and is the focus of this work. In the following we describe a method for revealing topological information of covariates and show how it can be used to systematically identify which are the relevant ones and provide indications as to what possible unknown covariates might be.

Our motivation comes from recent applications of topological methods—those central to the field of topological data analysis [7–10]—to the study of information encoded by neurons [11]. In particular, several authors have studied the topological aspects of the covariates of neuronal activity [12–15]. As an example of what is meant by topological aspects of the covariates of neuronal activity, Dabaghian *et al.* [13] consider the activity of a simulated population of place cells—neurons in rats and other mammals that preferentially fire when the animal is located in certain regions of its surrounding space [16]—and use persistent homology to recover the homology of the physical space. They are thus able, for example, to count the number of obstacles in the space explored by the simulated animal using only the spike trains of neuronal firing as input. Giusti *et al.* [15] use persistent homology of certain simplicial complexes built from correlations of neural recordings to show that the neuron population under observation encodes geometry. We can see here that what is meant by the state space of the activity is a combination of the selectivity of the neurons, the properties of the experiment, and other aspects of the system modeled (e.g., the behavior of an animal).

These topological methods, therefore, are useful for identifying the geometry of state space, given data. On the other hand the statistical models can be constructed to include the identified covariates and testing their relevance. In this work we have sought to bring together these two methods such that the topological insight can guide the construction of the statistical model, which can then be used to generate residual data that might contain additional information about any unknown features still remaining in

*gard.spreemann@epfl.ch; Present address: Laboratory for Topology and Neuroscience, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland.

†Present address: Lehrstuhl für Geometrie und Visualisierung, Technische Universität München, 85748 Garching bei München, Germany.

the data, or feed back into the topological side for further analysis.

For simplicity, we work with two-state model neurons whose interactions are given by an arbitrary graph and whose dynamics are governed by a Bernoulli GLM or kinetic Ising model with multiple external fields. We assume that the domain of these external fields is some reasonably well-behaved state space manifold, and it is this manifold that we seek to reveal topological information about by observing the dynamics of a realization of the system. We refer to [17] for different neuron models in the context of studying neuronal systems using homology.

We do not primarily seek to recover the homology of the state space itself (indeed, the experiments in Secs. IV B and IV D illustrate that that is not required in our method), although this is also discussed and is a part of the method. Rather, the questions we address in this paper are the following:

(i) Given data from a neural recording, and supposing that we know some external covariates contributing to the observed data, how do we decide whether the known covariates account for the data?

(ii) Are there other hidden covariates to be revealed?

(iii) If so, can topology help to reveal information about these?

Very briefly summarized, our method starts by building a certain filtered simplicial complex (a combinatorially constructed space generalizing the notion of a graph) from correlations of observations of the binary dynamics of the nodes in the system. Under some assumptions, the correlations of neuron activity are proxies of intersection information for regions of elevated activity in the state space, and the simplicial complex we build approximates the nerve of those regions, as illustrated in Fig. 1. By inferring away the contributions of known covariates from the observations, we reveal geometric information about any remaining, unknown covariates by computing persistent homology of a filtered simplicial complex built from the observations.

Finally, we point out that our method should also be applicable to the large class of real systems whose behavior is sufficiently well described by a kinetic Ising model. Examples of such are (real) neural recordings [18], stock markets [19], and contagion dynamics [20].

This project grew out of work in the first author's thesis, and a much lengthier and substantially different presentation can be found in the preprint [21].

II. BACKGROUND

We briefly recap the kinetic Ising model and the prerequisite constructions from topology and topological data analysis. See for example [22] for an introduction to the former, and the textbooks [9,10,23] or the survey article [7] for the latter.

A. Kinetic Ising model

The kinetic Ising model governs a set of ± 1-valued random variables $S_1(k), \ldots, S_N(k)$ at time steps $k = 1, 2, \ldots$. The model is specified by couplings $J \in \mathbb{R}^{N \times N}$, which define the neighbor relationships and their strengths, and

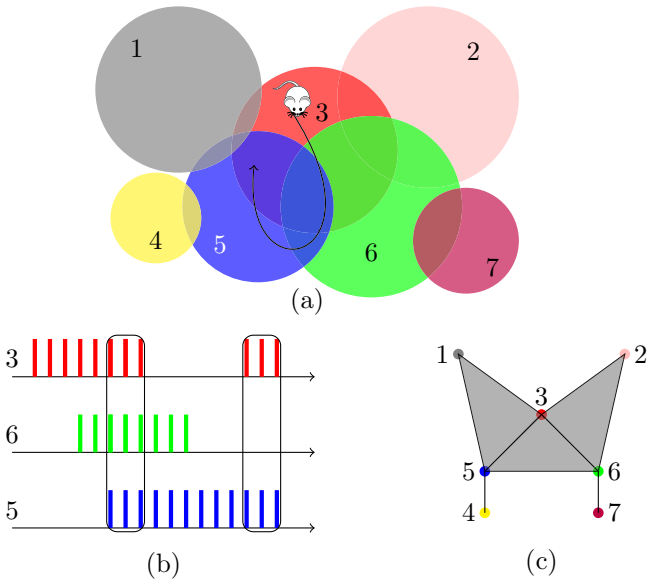

FIG. 1. Neuron cofiring is a proxy for firing field intersections, here illustrated in the motivational neuroscience application in a purely spatial setting. As the animal moves along the indicated path (a) in state space (here the same as physical space), we might observe the (highly idealized) firing events (b) of the corresponding place cells (nodes in the kinetic Ising model). The firing events in the leftmost box are indicative of the triple intersection of the red (3), green (6), and blue (5) place fields (regions of elevated activity), and those in the rightmost box are indicative of the double intersection of the red (3) and blue (5) regions. After the space has been thoroughly explored, the intersection data obtained from the cofiring of the spike trains are translated into a space (c) by means of an approximation to the nerve construction, as described in Sec. II B. This space and the space covered by place fields, i.e., physical space in this example and the abstract state space in general, share some topological properties that are reflected in their persistent homology.

external fields $E_1, \ldots, E_N : \mathbb{N} \to \mathbb{R}$ which define the influence of external covariates on the system as a function of time steps. We shall later assume that the external fields are in fact functions of the system's fixed state space, and that the temporal dependence comes from precomposing with the system's evolution path. If each variable $S_i(l)$ took the value $s_i(l) \in \{-1, +1\}$ for $l \leqslant k$, then at time step $k+1$, the conditional probability of observing $+1$ from the variable $S_i(k+1)$ (that is, that "neuron i fires" in the neuroscience interpretation) is

$$P[S_i(k+1) = 1 \mid S_1(k) = s_1(k), \ldots, S_N(k) = s_N(k)] = \frac{\exp F_i(k)}{2 \cosh F_i(k)},$$

where $F_i(k) = E_i(k) + \sum_{j=1}^{N} J_{i,j} s_j(k)$. For ease of notation, we set $s_1(0) = \cdots = s_N(0) = -1$. This random initial state should not significantly affect our method, as such initial noise will be no more influential than spontaneous firing at later times.

We refer to the expression $F_i(k)$ as the system's *Hamiltonian* (at time step k).

We assume that the temporal variation of the external fields E_i arise in a quite particular way through a random walk

on, or other exploration of, the state space M. The precise definition is given in Sec. III. For now it suffices to know that the relationship between the external fields, the state space, and the random walk are such that there are regions of state space where a certain node is more likely to be in state $+1$. Adopting the language of neuroscience, the region of elevated activity for a node will be referred to as the node's *place field*. Hopefully the two unrelated uses of "field" in "place field" and "external field" do not cause confusion.

B. From binary dynamics to spaces

Our method is topological in nature, and we will therefore construct certain topological spaces from observations of the system under study. These spaces, called *simplicial complexes*, can be thought of as higher-dimensional generalizations of graphs, or discrete versions of topological spaces. Abstractly, simplicial complexes consist of pieces of various dimension: vertices (or zero-simplices), edges (one-simplices), filled triangles (two-simplices), filled tetrahedra (three-simplices), and so on. Analogously with how an edge in a graph is associated to precisely two distinct vertices, p-simplices in a simplicial complex are associated to $p+1$ distinct $(p-1)$-simplices, called its *faces*, also required to be present in the complex. The lower right part of Figure 1 shows an example of a simplicial complex. It has seven zero-simplices, nine one-simplices, and three two-simplices.

Though defined abstractly, we often think of a simplicial complex K geometrically as above, which is to say we consider its *geometric realization*, a topological space $|K|$.

One interesting question that can be asked of a graph is the number of cycles it has, i.e., its number of holes. The analog of this property for a simplicial complex is much more interesting, for there *cycle* and *hole* are distinct notions; the latter being the former without triangles filling it in. Moreover, one has the higher-dimensional analogs of cycles and holes built from higher-dimensional simplices rather than edges. In general, a p cycle will be a collection of p-dimensional simplices having no boundary, and a $(p+1)$-dimensional hole will be a p cycle not filled in by a collection of $(p+1)$-dimensional simplices. *Homology*, the definition of which is recalled in Sec. II C, captures this notion precisely with linear algebra. It will be the tool we use to analyze simplicial complexes built from observations of the system in a manner that we now describe. Throughout the present section, we simply let homology be some vector space associated to a simplicial complex or to a topological space that is assumed to provide useful information about said complex or space.

Our method is concerned with revealing information about the system's *a priori* unknown state space, so we will obviously not be attempting to build that space as a simplicial complex. At the heart of our method is the *nerve construction*.

Definition 1. Let $\mathcal{U} = \{U_i \mid i \in I\}$ be a finite collection of sets. The *nerve of* $\mathcal{U}$ is a simplicial complex $N\mathcal{U}$ that contains a p-simplex $[i_0, \ldots, i_p]$, $p \geqslant 0$, if and only if $U_{i_0} \cap U_{i_1} \cap \cdots \cap U_{i_p} \neq \varnothing$.

The nerve of a collection of sets is a simplicial complex that has a vertex for each of the sets, an edge between any pair of vertices representing overlapping sets, a triangle between any triple of vertices representing overlapping sets, and so forth.

Under some assumption of the underlying set of sets, the nerve contains a lot of information. The *nerve theorem* ensures that if a collection of sets is a nice enough cover of a space X, then (the geometric realization of) the nerve of that collection is a topological space that can be continuously deformed into X itself without tearing. The precise statement follows.[1]

Theorem 2. Let X be a metric space, and let $\mathcal{U} = \{U_i \mid i \in I\}$ be a finite cover of X by closed sets with the property that for all subsets $J \subseteq I$, the intersection $\bigcap_{j\in J} U_j$ is either empty or contractible. Then X and $|N\mathcal{U}|$ have the same homotopy type.

(For a proof, see Corollary 4G.3 of [24].) In particular, it follows that a space and the nerve of a cover satisfying the above conditions have the same homology. This is crucial to our method because the nerve is a purely combinatorial construction.

We think of the nodes in the Ising model as each having a region U_i in state space X wherein its activity is elevated (place fields in the neuroscience setting). Even though X and each U_i are unknown, we can get some information about *intersections* of various U_i's by observing the correlations of the activities of the corresponding nodes in the system. Figure 1 illustrates the idea in the neuroscience setting: if n place fields intersect, then the corresponding n neurons are likely to fire together often. Cofiring of neurons is thus a proxy for the intersection information of their fields of elevated activity, and we can go from observing *binary time series or spike trains*, to *intersection information*, to a *simplicial complex*. If we assume that the U_i's are nice enough to satisfy the conditions of Theorem 2, we are now guaranteed that homology of $N(\{U_i\}_i)$ is the same as that of X.

The overlapping or not of regions of elevated activity in state space should, however, not be considered as a binary property. Just as correlation is a continuous measure, so is degree of overlap. A set of nodes whose observed dynamics are highly correlated are likely to have regions of elevated activity in state space that overlap greatly, such as the red, green, and blue neurons in Fig. 1. Similarly, a set of uncorrelated nodes correspond to regions of elevated activity that do not overlap. We have thus moved from a binary measure of intersection to a range from "not overlapping" to "highly overlapping." This is suitable for analysis with *persistent homology*, to be detailed in Sec. II C.

Input for persistent homology takes the form of a *filtered* simplicial complex, which is a sequence of simplicial complexes K_ε such that $K_\varepsilon \subseteq K_\delta$ whenever $\varepsilon \leqslant \delta$. We refer to the simplicial complex K_ε as that of the filtration at *scale*ε. Correlations (or other real-valued functions on pairs) give rise to filtered simplicial complexes as follows.

As part of our input (the rest is described in Sec. III), we take observations of N nodes in the kinetic Ising model over T time steps as binary vectors $s_i = (s_i(1), \ldots, s_i(T)) \in \{-1,1\}^T$, $i = 1, \ldots, N$. For each pair of observations we compute the

[1] A reader unfamiliar with the notion of "homotopy type" may think of two spaces as having the same homotopy type if and only if they can be continuously deformed into one another without tearing. See for example [24].

Pearson correlation[2] $\mathrm{corr}(s,s')$. The computations that follow could be based on these correlations directly, but as was argued in [15], one may instead want to consider only the *order* of the correlations. For $i,j \in \{1,\ldots,N\}$, we thus sort the values of $1-\mathrm{corr}(s_i,s_j)$ in ascending order, breaking ties arbitrarily. In the complete graph on the vertices $1,\ldots,N$, we assign a weight $w(i,j)$ to the edge (i,j) according to the corresponding position in the above ordering, normalized so that the edge corresponding to the two nodes with lowest correlation is assigned a value of 1.

Starting from the weighted graph G coming from correlation order as described above, we build the *flag complex* (or *Rips complex*) $R(G)$: its zero- and one-simplices are the vertices and edges of G, respectively, weighted by 0 and the corresponding edge weights; for each three-clique in the graph, add a two-simplex with weight equal to the maximum of the weights of the three edges in the clique; for each four-clique in the graph, add a three-simplex with weight equal to the maximum of the weights of the four two-simplices in the clique; and so forth. A filtration of $R(G)$ is given by taking as $R(G)_\varepsilon$ all the simplices with weight $\leqslant \varepsilon$.

C. Persistent homology

For an introduction to homology in general, both the notion for topological spaces and that for simplicial complexes, see [24]. Here we describe the simplicial notion tersely.

Let K be a simplicial complex. Algebraically, we think of collections of p-simplices, called *p-chains*, as finite formal sums of p-simplices with coefficients 0 or 1. Under addition modulo 2, this is a $\mathbb{Z}/2\mathbb{Z}$-vector space $C_p(K)$ with the p-chains of K as its basis. For notational reasons, we define C_{-1} to be the trivial vector space.

For all $p \geqslant 0$ there is a linear function $\delta_p : C_p(K) \to C_{p-1}(K)$ defined on basis elements of $C_p(K)$ by sending a p-simplex to the sum of $(p-1)$-simplices that are its faces. One verifies that $\delta_{p-1} \circ \delta_p = 0$. Chains in $\ker \delta_p$ are called *p-cycles* for reasons that become geometrically apparent from Fig. 2. Chains in $\mathrm{im}\,\delta_p$ are likewise called *$(p+1)$-boundaries*.

As discussed in Sec. II B, homology measures cycles that are not filled in, i.e., cycles that are not boundaries. Thus, one defines the pth homology of K as the quotient vector space $H_p(K) = \ker \delta_p / \mathrm{im}\,\delta_{p+1}$. Figure 2 shows a geometric interpretation of cycles in a homology class.

Persistent homology takes a filtration as input and tracks homology across all scales. $\mathcal{H}_p(K)$ consists of all the vector spaces $H_p(K_\varepsilon)$ for each ε, as well as the linear functions $H_p(K_\varepsilon) \to H_p(K_\delta)$ coming from the inclusions $K_\varepsilon \subseteq K_\delta$ whenever $\varepsilon \leqslant \delta$. These linear functions then track the lifespan of homology classes.

Theorems from algebra [25,26] ensure that $\mathcal{H}_p(K)$ can be fully specified in a compact manageable way, through its so-called *barcode decomposition*. Because of this decomposition,

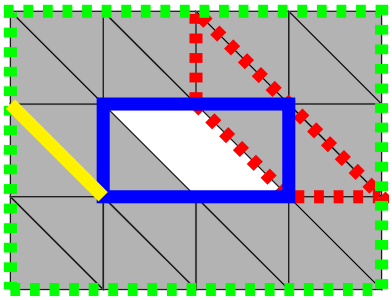

FIG. 2. A simplicial complex K with 20 zero-simplices, 38 one-simplices, and 22 two-simplices, and some highlighted one-chains. The solid yellow (light) one-chain consists of a single one-simplex, and is neither a cycle nor a boundary. The dashed red (dark) one-chain has trivial boundary, and is therefore a cycle. It is not a representative of any nontrivial homology class, for it is the boundary of two-chain consisting of the three two-simplices it encloses. The dashed green (light) and the solid blue (dark) one-chains are cycles that represent the same homology class (intuitively the two-dimensional hole in the middle). $H_0(K)$ is one dimensional, reflecting K's single connected component, while $H_1(K)$ is one dimensional due to the central hole.

persistent homology of a filtered simplicial complex can be represented and drawn as a *persistence diagram*, a plot of points (with multiplicity) above the diagonal in $(\mathbb{R} \cup \{\infty\})^2$. The horizontal coordinate indicates the *birth scale* of some homology class and vertical coordinate indicates its *death scale*. Figure 3 shows an example.

When multisets of points in the extended plane are inconvenient to work with, one may also discard some information and simply keep track of the *number* of homology classes present

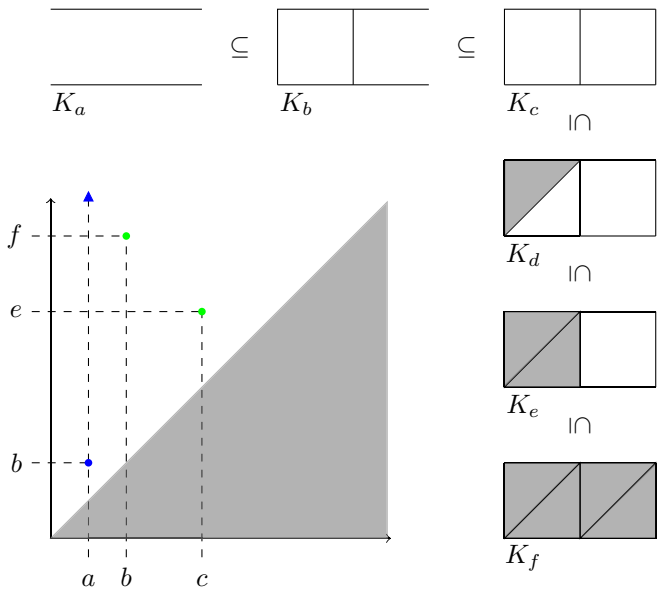

FIG. 3. On the outer rim is shown a filtration $K = (K_a, K_b, K_c, K_d, K_e, K_f)$ for some $0 < a < b < c < d < e < f$. The persistence diagram shown is an overlay of $\mathcal{H}_0(K)$ in blue (dark) and $\mathcal{H}_1(K)$ in green (light). The triangle symbolizes the infinite summand of $\mathcal{H}_0(K)$, i.e., the single connected component of K that cannot disappear. Both classes in $\mathcal{H}_1(K)$, i.e., two-dimensional holes, die at finite scales (e and f).

[2]When working with data collected from real physical experiments it may be useful to replace the Pearson correlation with an average of Pearson correlations of *shifted* observations in order for the method to be less vulnerable to timing or desynchronization errors in observations.

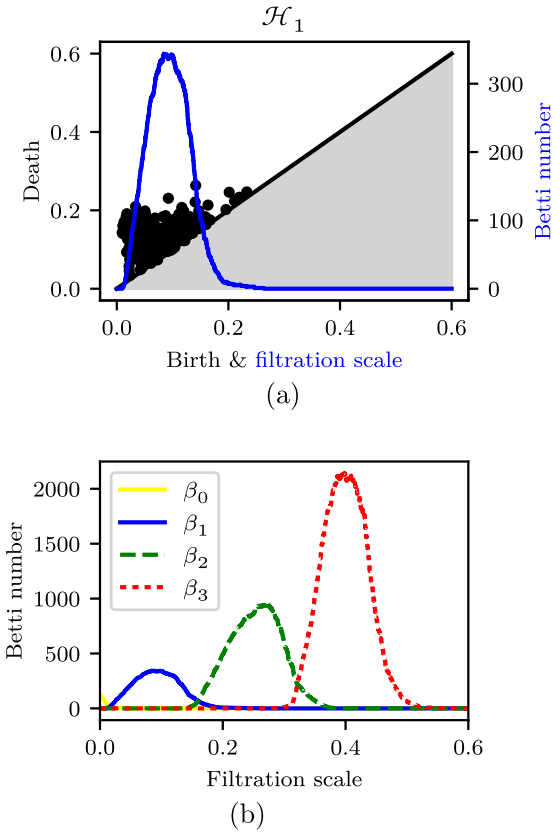

FIG. 4. Persistent homology of a realization of an Erdős-Rényi complex with 120 vertices. Increasing homology dimension results in an increase in peak Betti number and a narrowing of the edge probability range where nonvanishing Betti numbers are likely. This behavior is a signature of ER complexes. (a) Persistence diagram and Betti curve for $\mathcal{H}_1$. (b) Betti curves for $\mathcal{H}_0$, $\mathcal{H}_1$, $\mathcal{H}_2$, and $\mathcal{H}_3$.

at each scale and thus reduce a persistence diagram to a *Betti curve*. An example of this is shown in Fig. 4.

D. Statistics and comparison of persistence diagrams

When we later infer away covariates of a system, we need to compare filtered simplicial complexes to those arising from certain data models. One important such model is that of random data. In this section we give two *ad hoc* measures of closeness to a random system and argue for their validity. Comparison with filtered simplicial complexes arising from other models will depend on the model. In particular, one may look to [15] for comparison with complexes arising from geometric data.

The survey by Kahle [27] covers much of what is known about the homology of some special cases of random simplicial complexes, namely those that are flag complexes of Erdős-Rényi (ER) random graphs. An (n,p)–*Erdős-Rényi* random graph has n nodes, and each possible edge appears independently with probability p. We will refer to flag complexes of such graphs as *ER complexes* with parameters n and p. While most results known are asymptotic in the number of vertices in the complex, and thus of little direct relevance to our setting, some qualitative conclusions can be drawn also about the finite case assuming the size of the vertex set is not too small.

The first qualitative observation is that homology is likely to occur only in ER complexes when the edge probability parameter is in a certain range, and this range becomes narrower as the homology dimension grows. Within the "allowed range," however, a large Betti number may occur. As the homology dimension increases, the allowed range shrinks while the peak Betti number within the range grows. Figure 4 illustrates the behavior. We will use this behavior as a signature of the underlying graph having random weights. The persistent homology of our correlations complexes will be compared with these random ones in two different ways, which we now detail.

If G is a complete weighted graph on a vertex set V, we write $S(G)$ for (a realization of) the complete graph on V having the edge weights of G randomly shuffled. Except in degenerate cases $R(S(G))$ should be a good realization of the order flag complex of an ER random graph. The discussion above then suggests two *ad hoc* measures for how consistent a flag complex is with an ER random complex by comparing $\mathcal{H}_*(R(G))$ and $\mathcal{H}_*(R(S(G)))$.

For various p and q we define

$$\delta_k(G) = d_{p,q}(\mathcal{H}_k(R(G)),\mathcal{H}_k(R(S(G)))),$$

where $d_{p,q}$ is the (p,q)-persistence module metric.[3] We further let β_k and β'_k denote the Betti curve of $\mathcal{H}_k(R(G))$ and of $\mathcal{H}_k(R(S(G)))$, respectively, and define the ratio

$$\Delta_k(G) = \frac{\max_s \beta_k(s)}{\max_s \beta'_k(s)}$$

whenever it exists. $\Delta_k(G)$ thus compares the peak Betti number of $R(G)$ to that of its shuffled version. If, for as large a k as is computationally feasible, $\delta_k(G)$ is close to 0 and $\Delta_k(G)$ is close to 1, we have an indication that $R(G)$ resembles an ER random complex.

III. UNCOVERING HIDDEN COVARIATES

We suppose that some, but not all, covariates of the activity of the nodes in the system are known. Measurements of these known covariates throughout the observation of the system supplement the time series of node dynamics observations as our input data. We further suppose that the known covariates can be described by certain functions of distances on some simple manifolds whose product makes up state space. We will specifically consider the following:

(i) boxes $(0,1)^d$ of any dimension d, with the Euclidean metric,

(ii) boxes with any number number of d disks removed so long as these do not disconnect the manifold,

(iii) circles, with their usual metric,

(iv) spheres, with their usual metric,

(v) products of the above, with the product metric.

[3]The (p,q)-persistence module metric between two persistence diagrams X and Y is $d_{p,q}(X,Y) = \min_f [\sum_{x\in X'} \|x - f(x)\|_p^q]^{1/q}$, where f ranges over all multiset bijections $f : X' \to Y'$, X' and Y' are X and Y with infinite multiplicity everywhere on the diagonals (so bijections exist), and $\|\cdot\|_p$ denotes the Euclidean p-norm.

Section V discusses a generalization, and the fact that in some of these cases (such as boxes), edge effects may introduce problems.

The true state space of the system is assumed to be a product $M = M_1 \times \cdots \times M_L$ with each factor a manifold from the list above corresponding to a covariate. If only $P < L$ of the covariates are known, we can only observe a partial state space $\widetilde{M}$. For ease of notation we will assume that the factors $M_1, \ldots, M_P$ are the known ones.

The system is observed by recording the activity of its nodes, together with samples of the trajectory of the system through the known factors of state space at T time steps. Our measurement data thus consist of N binary vectors of the form

$$s_i = (s_i(1), s_i(2), \ldots, s_i(T)) \in \{-1,1\}^T,$$

together with samples of the observable coordinates in the state space. If we write $\alpha : \mathbb{R} \to M$ for the temporal evolution of the system through state space, we record samples of this path as

$$\widetilde{\pi}(\alpha(1)), \widetilde{\pi}(\alpha(2)), \ldots, \widetilde{\pi}(\alpha(T)) \in \widetilde{M}.$$

Here $\widetilde{\pi} : M \to \widetilde{M}$ denotes projection onto the known factors of M.

We now assume that the external field part of the Hamiltonian decomposes into a sum of Gaussians on the various factors of the state space. Corresponding to known covariate number $1 \leqslant l \leqslant P$ is a linear combination of Q_l Gaussians, each having parameters $\sigma_{l,q} \in \mathbb{R}$ and $c_{l,q} \in M_l$. Projections onto individual factors (both known and unknown) are written $\pi_l : M \to M_l$, and we denote the metric on M_l by d_l. Define Gaussians $V_{l,q} : M_l \to \mathbb{R}$ by

$$V_{l,q}(x) = \exp\left(-\frac{(d_l(x, c_{l,q}))^2}{2\sigma_{l,q}^2}\right)$$

for $1 \leqslant l \leqslant P$ and $1 \leqslant q \leqslant Q_l$, and assume that the external part of the Hamiltonian for node i at time step k can be written (recall that α is the system's path through its state space)

$$E_i(k) = \sum_{l=1}^{L} \sum_{q=1}^{Q_l} A_{i,l,q}(V_{l,q} \circ \pi_l \circ \alpha)(k).$$

To gain some intuition about the above expressions, it may be instructive to consider the special case where for each i and l there is only one q for which $A_{i,l,q} \neq 0$. This is also the setting in which all of our data for the experiments in Sec. IV are generated. Fix an i and an l and let q be the only index for which $A_{i,l,q} \neq 0$. Then $c_{l,q}$ is the center of a region of elevated activity corresponding to covariate number l, $\sigma_{l,q}$ is a measure of its width, while $A_{i,l,q}$ specifies the peak strength with which it influences node i in the system (the sign determining whether the influence is activating or suppressing). In the neuroscience example, neuron i has a place field centered at $c_{l,q}$ with size controlled by $\sigma_{l,q}$. When the above relationship between the indices of $A_{\bullet,l,\bullet}$ does not hold, we allow each covariate to govern the activity of nodes in the model through a linear combination of Gaussian fields, which will be necessary for the inference process described next.

The likelihood of data,

$$\mathbf{s} = \{s_i(k) \mid 1 \leqslant i \leqslant N, 1 \leqslant k \leqslant T\},$$

observed under the kinetic Ising model is

$$L(\mathbf{s}) = \prod_{i=1}^{N} \prod_{k=1}^{T-1} \frac{\exp[s_i(k+1)F_i(k)]}{2\cosh F_i(k)}.$$

Its logarithm is convex in $A_{\bullet,\bullet,\bullet}$ and $J_{\bullet,\bullet}$, and we can therefore perform likelihood maximization by means of convex optimization to infer the values of these parameters that best fit the observed data. Should such optimization be too computationally expensive, one may employ heuristic or approximate methods instead, though that was not a problem encountered in our experiments. With these coefficients in hand, we can selectively subtract from the spike trains the (expected) contribution of the candidate covariates as provided by the model with the inferred coefficients. The residual data after removal no longer consist of binary observations, but instead are real-valued discrete time series for each node. The interpretation of such residuals is fraught with difficulties, but as we will only consider their correlations we do not touch upon this problem.

To illustrate our method (in Sec. IV), we will work with spatial and head direction tuning in the context of the motivational neuroscience example. Throughout, all the inference will be performed with $25^2 = 625$ spatial basis functions, and 25 circular basis functions, both with uniformly distributed means. For example, if the known factors of state space are $\widetilde{M} = M_1 \times M_2$ with $M_1 = (0,1)^2$ and $M_2 = \mathbb{S}^1$, then $Q_1 = 25^2$ and $Q_2 = 25$, and with the $c_{1,q}$'s forming a regular grid on $(0,1)^2$ and the $c_{2,q}$'s uniformly distributed on $\mathbb{S}^1$.

After we have removed the contributions from the known or suspected covariates, we build the flag complex of the graph weighted by correlations of residual (discrete) time series, and compute its persistent homology. The resulting persistence diagrams are then analyzed for conformity with known behavior of simplicial complexes arising from null-model networks. The fact that this is possibly to do with persistent homology has been demonstrated repeatedly [15,28,29]. For example if they are consistent with a random simplicial complex, as described in Sec. II D, we can be confident that there are no unknown covariates to the extent that our *ad hoc* measure of randomness holds up. If, on the other hand, they conform to data with a geometric structure, the work of Giusti *et al.* [15] shows that it is likely that further unaccounted-for covariates are present in the system. If in addition some of the geometric structure is discernible from the persistence diagrams, it will guide further model selection. In the neuroscience setting, such unaccounted-for covariates are frequently discovered [30], and their presence and homological properties may prove useful to their discovery.

The method can thus be written out as follows:

(i) Perform an experiment as described above, yielding binary time series $s_1, \ldots, s_N$ and samples of (some of the) known coordinates of the state space path $\widetilde{\pi} \circ \alpha$.

(ii) Compute the correlation order weights $w(i,j)$ for the complete graph G on N vertices.

(iii) Compute $\mathcal{H}_*(R(G))$.

(iv) If $\mathcal{H}_*(R(G))$ is consistent with the persistent homology of a filtered simplicial complex coming from a null-model network (e.g., ER random complex per Sec. II D), go to step

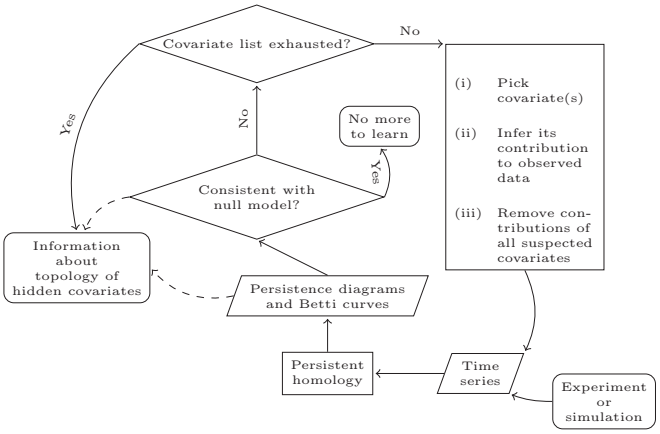

FIG. 5. Our main contribution summarized. The process starts in the lower right corner with input data. At every iteration of the loop one may be able to learn new properties of the system (dashed lines) until the candidate covariate list has been exhausted, at which point topological properties of hidden covariates are extracted.

(iv)(a). If it is consistent with some other known complex structure, such as the geometric one of [15], go to step III.

(a) There is nothing more to learn from our method. We are done.

(v) Is the list of candidate stimuli exhausted? If so, go to step (v)(a). If not, go to step (vi).

(a) There are likely further covariates of a geometric or otherwise interesting nature. Moreover, $\mathcal{H}_*(R(G))$ may reveal information about the homology of the state space corresponding to unknown covariate(s).

(vi) Pick another suspected covariate (if possible, one informed by the persistence diagrams), and use the inference process described above to remove the contribution of this and all previously chosen covariates. This yields real-valued residual time series that from now on replace the binary ones $s_1, \ldots, s_N$. Go to step (ii).

Figure 5 summarizes the above procedure.

We finally point out that it is *not* essential to our method that we necessarily capture the correct homology of the full state space. It is the *presence of geometric or otherwise interesting structure* incompatible with null-model networks, and any possible homological feature of that, that is essential.

We end this section with an illustration of our method in a concrete setting. Again we return to the motivational example from neuroscience where nodes in the kinetic Ising model are neurons, and these are influenced by two simple covariates: the spatial location of the animal as it explores a box $B = (0,1)^2$, and the animal's head direction parametrized by the circle $\mathbb{S}^1$. Let us imagine that only the first is known. The full state space is therefore $M = B \times \mathbb{S}^1$, and the known covariate space corresponds to $\widetilde{M} = B$. An experiment is performed wherein the activity of N neurons are recorded at the same time as the animal's position in B (the only known factor of the state space). Supposing that $N = 100$ neurons are recorded over the course of 10 min,[4] the data we are given then consist of 100 binary spike train vectors of length 60 000 together with 60 000 samples of the animal state as points in the proposed state space $(0,1)^2$. Our goal is then to determine whether the suspected covariate—spatial position—describes the observed activity, and, most importantly, if not, what the homological properties of any remaining unknown stimuli are.

After having performed steps (ii)–(vi) once, the second iteration leads us to step (v)(a); there are no more stimuli to account for, yet we see persistent homology inconsistent with random data, and consistent with geometric data. Examination of the persistence diagrams of $\mathcal{H}_*(R(G))$ reveals that the unknown covariate(s) correspond to a state space with the same homology as a circle. Together with other evidence, this may lead the researchers to suspect head direction tuning as a hidden covariate. New experiments may then be performed to investigate this, and our method may be applied also to the new data.

IV. COMPUTATIONAL RESULTS

We test the efficacy of our method in some situations motivated by the neuroscience application. We use synthetically generated data in order to be in complete control of all "experimental parameters," and because publicly available real data often come from experiments without a topological focus (for example the spatial environment tends to be homologically trivial). For the results presented here, neuron activity was simulated from the same GLM as used for the inference process by appropriate selection of peak field strength coefficients $A_{\bullet,\bullet,\bullet}$, centers $c_{\bullet,\bullet}$ and widths $\sigma_{\bullet,\bullet}$. The field widths were selected to nicely cover state space for the given number of neurons, as with too few neurons or too small fields it is unreasonable to expect to be able to apply our method. In addition, a constant negative term (typically -1) was added to the external fields to make the overall firing rate low, but nonzero, outside of place fields, as is the case for many real cells. This term essentially just lowers the noise floor of our data, and should be of no deeper significance to us. All of the experiments except for those in Secs. IV C and IV E are performed on data generated from a coupling-free ($J_{\bullet,\bullet} = 0$) model.

When describing the experiments and their result, we will mostly, in accordance with our motivation, adopt the language of the neuroscience application of neurons or cells firing and tuning to biologically relevant covariates.

If the physical environment is denoted B, then the $B \times \mathbb{S}^1$ (factors of) state space is explored as follows: If at some time step the animal's state is $(x,y,\theta) \in B \times \mathbb{S}^1$, then the next state is found by choosing a θ' randomly and uniformly within 0.02 of θ in $\mathbb{S}^1$. If $(x + 5 \times 10^{-4} \cos\theta', y + 5 \times 10^{-4} \sin\theta', \theta')$ is within $B \times \mathbb{S}^1$, this point is the next state. If not, new angles are drawn until the new state is valid. Additional factors are explored similarly. We note that qualitatively similar results are obtained if this slightly realistic random walk is replaced by ordinary Brownian motion or uniform random sampling.

The experiments were all done with 100 neurons. See [17] for a discussion regarding the necessary number of neurons to recover one or multiple features. We point out that for hundreds of neurons, we do not face any obstacles of computational resources. Indeed, computations take on the order of minutes

[4]This is a realistically sized data set according to the computational neuroscience data sharing website https://crcns.org.

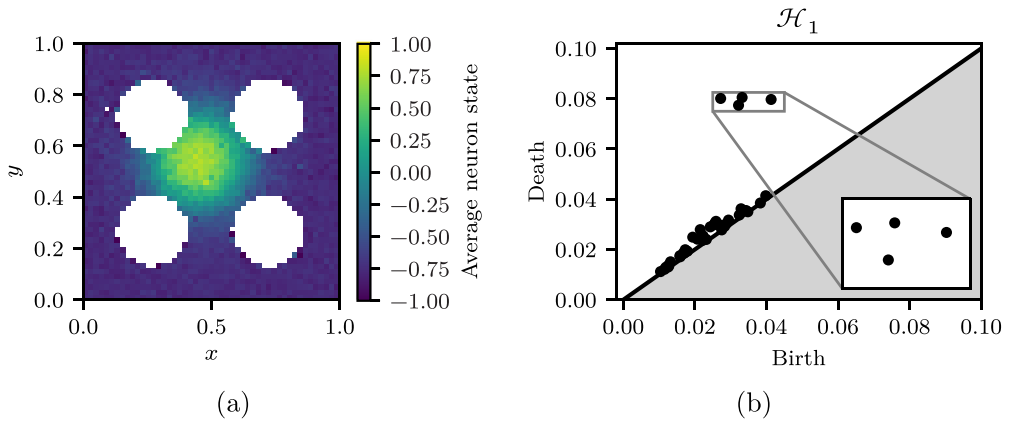

FIG. 6. Results of the experiment in Sec. IV A. (a) The activity of a single node/neuron as a function of spatial position. (b) Persistent homology of the space built from correlations. All higher dimensions of homology are trivial or close to trivial.

on moderate laptops (see also [31] for an overview of the real-world computational demands of persistent homology).

A. Recovering spatial homology

While the central point of this paper is the uncovering of the homological properties of unknown covariates, we begin by providing an example of how we can solve a simpler task: recovering the homology of a spatial environment. This is analogous to some of the results presented in [13].

State space is now the unit square punctured by four disks of radius 0.15. The disks are centered at (0.27,0.27), (0.27,0.72), (0.72,0.27), and (0.72,0.72). We refer to the punctured box as B below.

In the notation of Sec. III, $N = Q_1 = 100$, $M_1 = B$, $L = 1$, and $J_{i,j} = 0$ for all $1 \leqslant i, j \leqslant N$. The $c_{1,q}$'s form a regular grid on B, and

$$A_{i,1,q} = \begin{cases} 2 & \text{if } i = q \\ 0 & \text{otherwise.} \end{cases}$$

As an example, Fig. 6(a) shows the spatial tuning of a single neuron.

Figure 6(b) shows that we correctly recover the homology of B, the only part of M that influences neural activity. In other words, this example illustrates how we in the neuroscience setting can detect and even count the obstructions in a physical space from neuron activity alone.

B. Proof of concept

The simplest possible setting wherein our inference scheme, laid out in Sec. III, is useful, is perhaps one where both spatial and head direction tuning govern neuron activity, but where the researcher believes only one of those to be real.

For this computation, the spatial component of the state space is a unit square punctured in its center by a single disk of radius 0.2, denoted by B. The head direction component is $\mathbb{S}^1$.

In the notation of Sec. III, $N = Q_1 = Q_2 = 100$, $\widetilde{M} = M_1 = B$, $M_2 = \mathbb{S}^1$, $P = 1$, $L = 2$, and $J_{i,j} = 0$ for all $1 \leqslant i, j \leqslant N$. The $c_{1,q}$'s form a regular grid on B, and

$$A_{i,1,q} = \begin{cases} 2 & \text{if } i = q \\ 0 & \text{otherwise.} \end{cases}$$

Similarly, the $c_{2,q}$'s are uniformly spread out over $\mathbb{S}^1$. To avoid artificially coupling head direction and spatial tuning through the ordering of their place field centers, we let τ be a random permutation of $\{1, \ldots, N\}$ and then let

$$A_{i,2,q} = \begin{cases} 2 & \text{if } i = \tau(q) \\ 0 & \text{otherwise,} \end{cases}$$

for all i, q.

We reiterate that we assume that the researcher is *unaware of head direction tuning* as a real influence on place cell activity in this example; he believes spatial position is the only relevant stimulus. After conducting an experiment, he sees the neurons' spatial dependence exemplified in Fig. 7(a). The head direction dependence in Fig. 7(b) is *not* known to the researcher.

Persistent homology of the observed correlations can be seen in Figs. 7(c) and 7(d). Note that we do *not* observe the actual persistent homology of the state space (a thickened torus, having two features in $\mathcal{H}_1$ and one in $\mathcal{H}_2$). This is not entirely satisfactory, but at least the observed persistence diagrams indicate there is homologically nontrivial information present in the neuron activity, which should be sufficient motivation for the researcher to continue investigating.[5]

Satisfied that the persistence diagrams are consistent with his hypothesis about the relevant covariates, the researcher proceeds to the next step, namely removing the effect of the spatial covariate (the only one he is aware of) by the means described in Sec. III.

The resulting rate maps after removal on spatial activity tuning can be seen in Figs. 7(e) and 7(f). They are as expected, and obviously do not provide new information to the researcher. The persistence diagram in Fig. 7(g), however, shows that there is still homologically nontrivial information

[5]It turns out that by carefully balancing the relationship between the strength of the spatial and head direction tunings, one can obtain the correct persistent homology. This suggests that one of the torus radii simply collapses to a circle for many strength relationships.

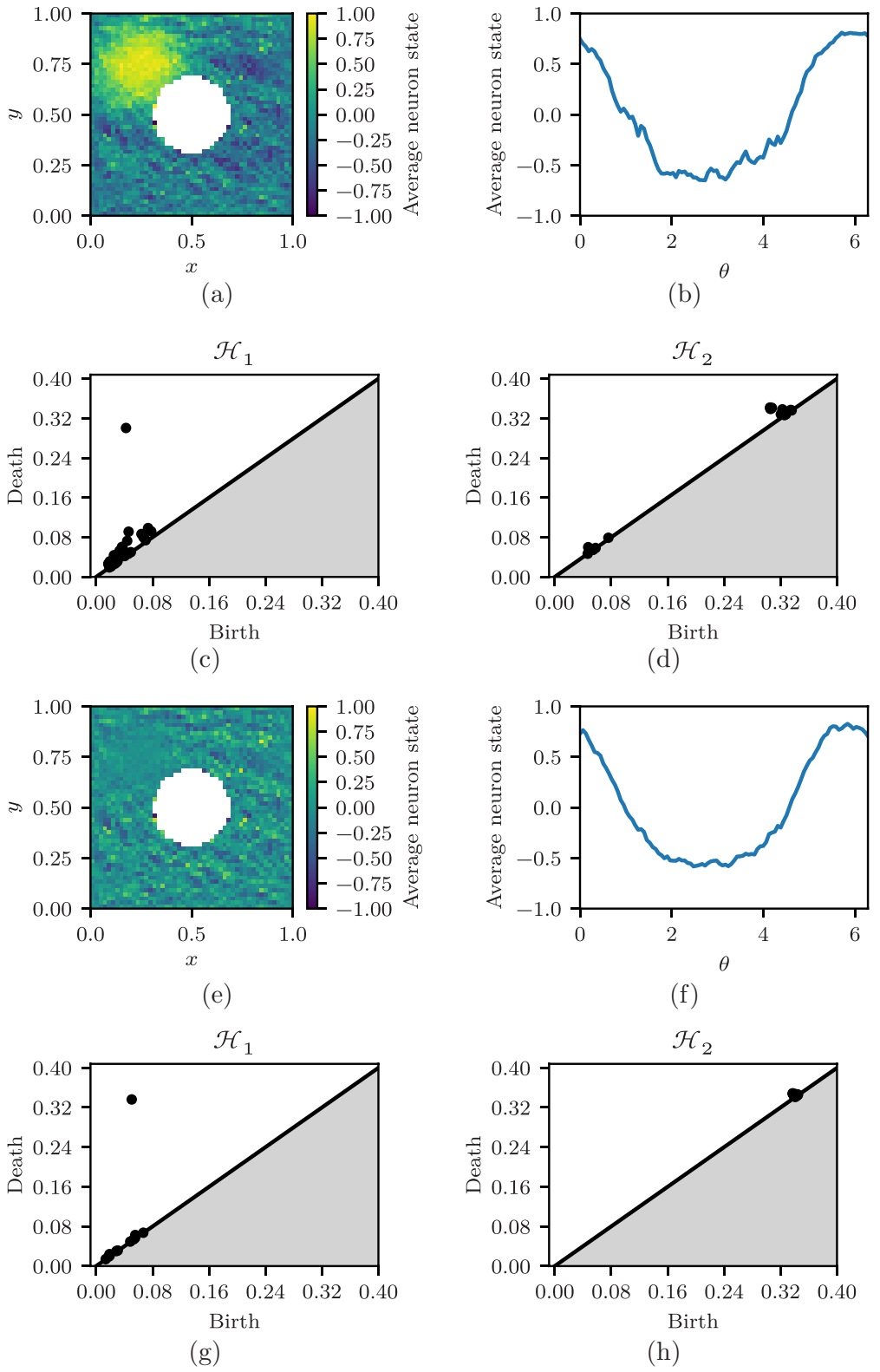

FIG. 7. Data and results throughout the experiment in Sec. IV B. (a) Activity of a single neuron as a function of spatial position. (b) Activity of a single neuron as a function of head direction. The researcher is *not privy to this information* in our experiment. (c),(d) Persistence diagrams of the space built from correlations at the start of the experiment before removing any contributions. (e),(f) Rate maps after removing influence from spatial position. (g),(h) Persistence diagrams after removing the influence of spatial position. A circular contribution remains!

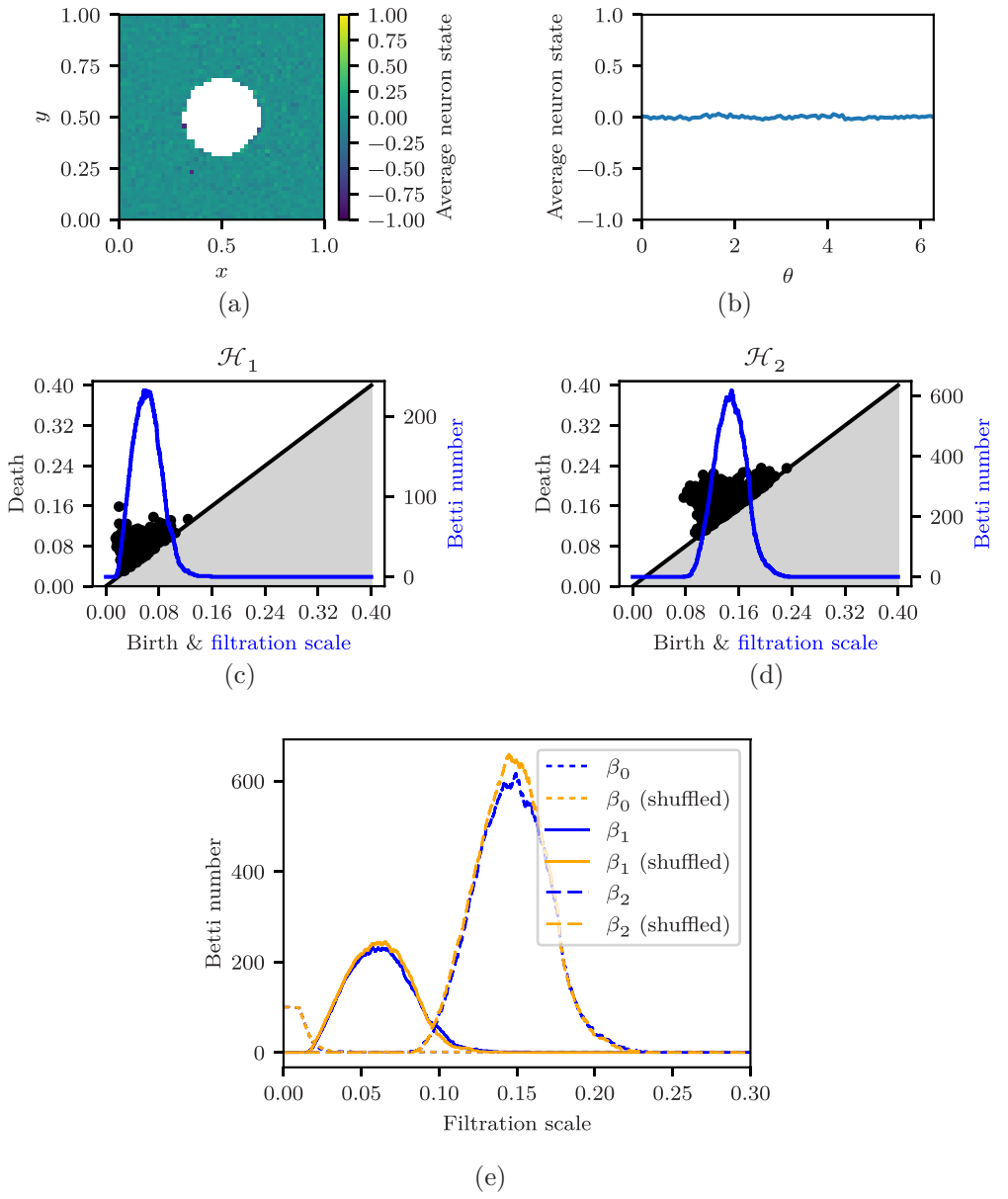

FIG. 8. Data and results throughout the experiment in Sec. IV B, after all contributions have been removed. (a) Activity of a single neuron as a function of spatial position. (b) Activity of a single neuron as a function of head direction. (c),(d) Persistence diagrams and Betti curves of the space built from correlations after all contributions have been removed. (e) Betti curves compared to those of typical ER random complexes, indicating that no structure is left. (The β_0 curves obscure each other.)

contained in the observed data. This should hopefully lead the researcher to suspect that there are further, hidden, influences on neuron activity, and, most importantly, that this or these *are of a circular nature*.

Guided by this, the researcher might consider head direction tuning. He therefore sets up a new experiment where also this is recorded, so that also its influence may be removed from the data. Doing so results in Fig. 8.

Alternative scenario

One may also want to consider an alternative hypothetical scenario wherein head direction tuning is the only suspected covariate. For readability reasons we do not include that scenario in full here. The interesting part is the persistent homology after the removal of head direction tuning. Figure 9 shows that we recover the correct $\mathcal{H}_1$ also in this case.

C. Effect of couplings

In the preceding experiments, neurons were never coupled. To illustrate that our method also copes with such "internal stimuli," we repeated the experiment from Sec. IV B with the change that every cell is given a weak but random coupling to a every other cell. Specifically, we kept all simulation parameters as before, but let every $J_{i,j}$ be drawn independently and uniformly from $[-0.1,0.1]$. The couplings are thus weak compared to the external stimuli (which peak at 2 in the centers of fields), but should nevertheless introduce noise to the data.

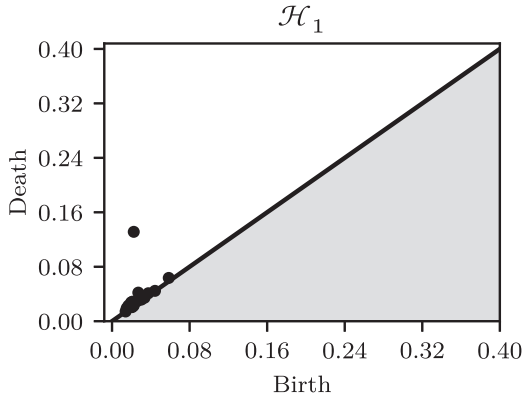

FIG. 9. Persistent homology after removing the effect of head direction tuning in an alternative version of the experiment in subsection of Sec. IV B.

Figures 10(a) and 10(b) show the results before any covariate removal. Observe that the random couplings introduce significant noise in the spatial dependence of the activity compared to that in Fig. 7(a). Figures 10(c) and 10(d) show that we are able to carry out the same procedure as in Sec. IV B also in the presence of couplings.

D. Effect of θ preference

Neurons tend to show increased activity at a certain phase of local oscillations, in particular in the range of θ. We simplistically model θ preference as each neuron preferentially firing near a randomly chosen phase of a 7 Hz sinusoidal wave in time. The experimental parameters are the same as in Sec. IV B, except now $L = 3$, and the state space gains an extra factor $M_3 = \mathbb{S}^1$. The $c_{3,q}$'s are uniformly spread out over $\mathbb{S}^1$, and for a random permutation[6] τ' of $\{1, \ldots, N\}$ the new field strength coefficients are

$$A_{i,3,q} = \begin{cases} 2 & \text{if } i = \tau'(q) \\ 0 & \text{otherwise,} \end{cases}$$

for all i,q.

θ preference is thus, as far as topology is concerned, precisely the same as head direction preference, except that the M_3 factor of state space is explored by always moving forward in time (modulo 1/14) instead of by a random walk. We therefore expect that θ preference will contribute to homology in the same way as head direction tuning. Figure 11(b) confirms this. Again it should be pointed out that we are not observing homology consistent with the three-dimensional torus $\mathbb{S}^1 \times \mathbb{S}^1 \times \mathbb{S}^1$. While this may seem unsatisfactory, it is a quite natural effect of one covariate suppressing the expression of the homology of the others, i.e., one of the radii of the torus dominating over the others. This illustrates well why the inference process and removal of covariates really is necessary;

[6]Present for the same reason as for the head direction tuning in Sec. IV B. The effect of more systematic couplings between related stimuli is investigated in [17].

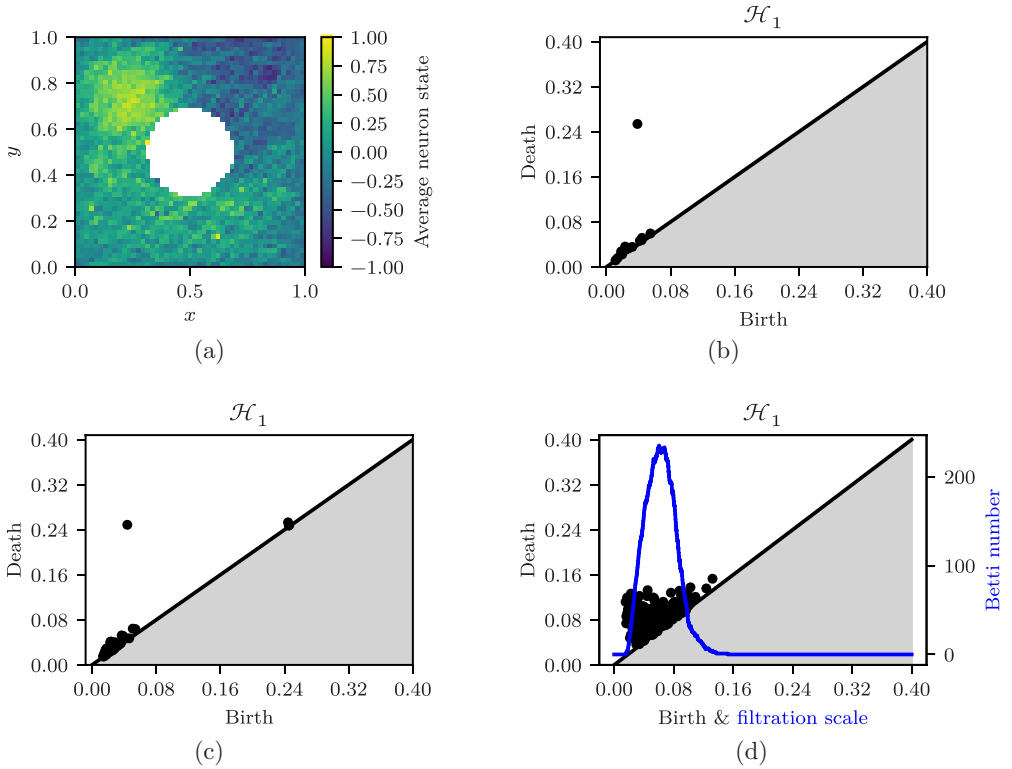

FIG. 10. Results from the experiment in Sec. IV C. (a) Activity of a single neuron as a function of spatial position before removing any covariates. (b) $\mathcal{H}_1$ before removing any covariates. (c) $\mathcal{H}_1$ after removing spatial dependence and internal couplings. A circular component remains! (d) $\mathcal{H}_1$ and Betti curve in dimension 1 after removing also head direction tuning. We see the characteristic signature of there being no more structure.

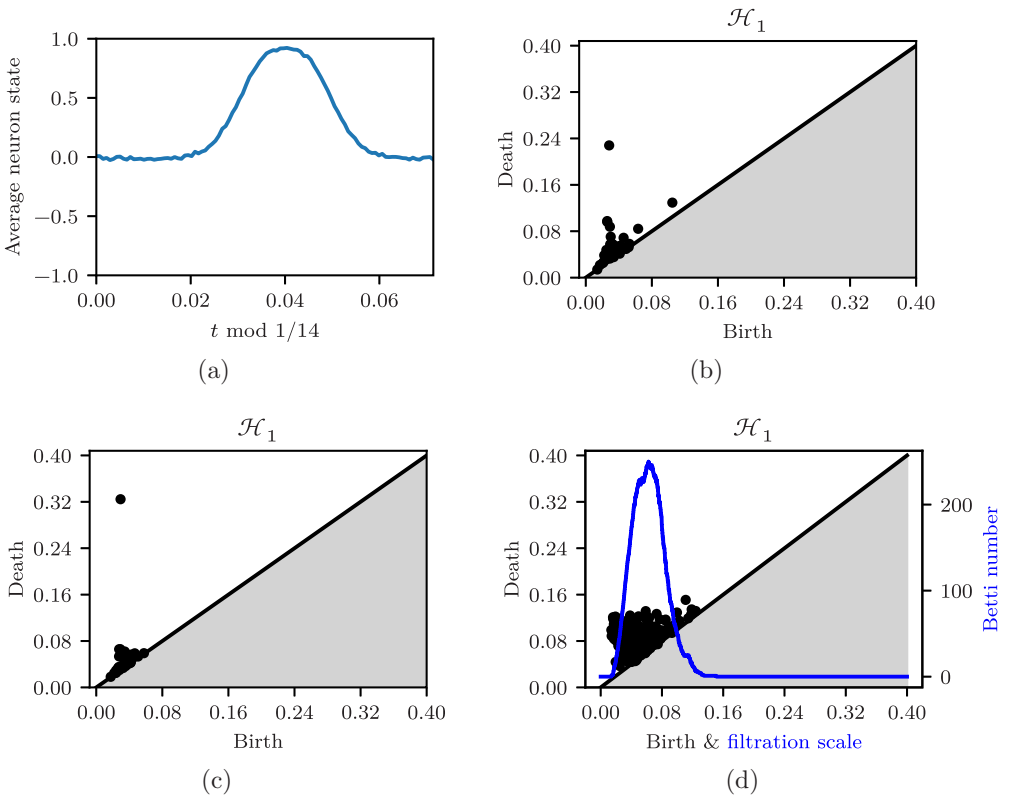

FIG. 11. Observations from the experiment in Sec. IV D. (a) Activity of a single neuron as a function of theta phase. (b) $\mathcal{H}_1$ before removing any covariates. (c) $\mathcal{H}_1$ after removing the contribution of both head direction and spatial position. A circular component remains! (d) After removing all covariates, we are left with a persistence diagram and Betti curve indicating that we are done.

the fact that there are *three* circular factors in the state space cannot be glanced at directly from the observed data.

Figures 11(c) and 11(d) show that we obtain the expected results when left with only θ preference and when all covariates are removed.

E. Homology from internal couplings

In the preceding experiments, internal couplings have been absent, or, as in Sec. IV C, not themselves been the focus of our attention. We now illustrate that our method is also capable of detecting the homology of (the flag complex of) the graph defining the neighbor relations of the neurons.

We generate data with only spatial fields and internal couplings. Specifically, $N = Q_1 = 100$, $M = M_1 = (0,1)^2$, the $c_{1,q}$'s form a regular grid on $(0,1)^2$, and the peak field strengths are

$$A_{i,1,q} = \begin{cases} 1 & \text{if } i = q \\ 0 & \text{otherwise.} \end{cases}$$

The symmetric matrix J describes a circle on all N nodes with edges in both directions with weights 2. The indices defining the edges of the circle are chosen randomly to avoid an unnatural coupling to the spatial fields through the ordering.

Figure 12(a) shows an unexpected result consistent with (at least) *two* circles. Manual inspection of the spike train distances reveals that this is an artifact of the GLM being coupled across only two consecutive time steps. When we use correlations of spike trains with no averaging over temporal shifts, we are unable to resolve any coupling interactions across an even number of neurons. The simplicial complex thus breaks into *two* circles, corresponding to the even and odd parity edges of the neighborhood graph $(\{1,\ldots,N\},\{(i,j) \mid J_{i,j} \neq 0 \text{ or } J_{j,i} \neq 0\})$. This undesired behavior vanishes when one replaces the correlation with an average of correlations of various *shifted* spike trains. For example, Figure 12(b) shows the case of averaging over ten shifts.

V. DISCUSSION

We believe that the core aspects of the method presented in this paper generalize to a wider setting. Imagine a point cloud $C = \{p_1, \ldots, p_N\}$ in some metric space (M,d) that we view as encoding the state of the system being observed. The exact further assumptions on M need to be worked out, for now we have only relied on numerical evidence. The pairwise distances of the points are *a priori* unknown. The system we measure consists of a Bernoulli process S_i, $1 \leqslant i \leqslant N$, for each point. A crucial assumption is that the "success" probability of process i at a time point is given by the distance from p_i to our state at that time. In other words, if we assume that we explore M through some time-dependent random walk $X(k)$, we assume at time step k that

$$P(S_i(k) = 1 \mid X(k) = x) = f(d(x, p_i)) \tag{1}$$

for some monotonically decreasing sufficiently integrable (unknown) function $f : \mathbb{R}^+ \to [0,1]$. If the random walk has progressed long enough that the distribution of $X(k)$ is close to

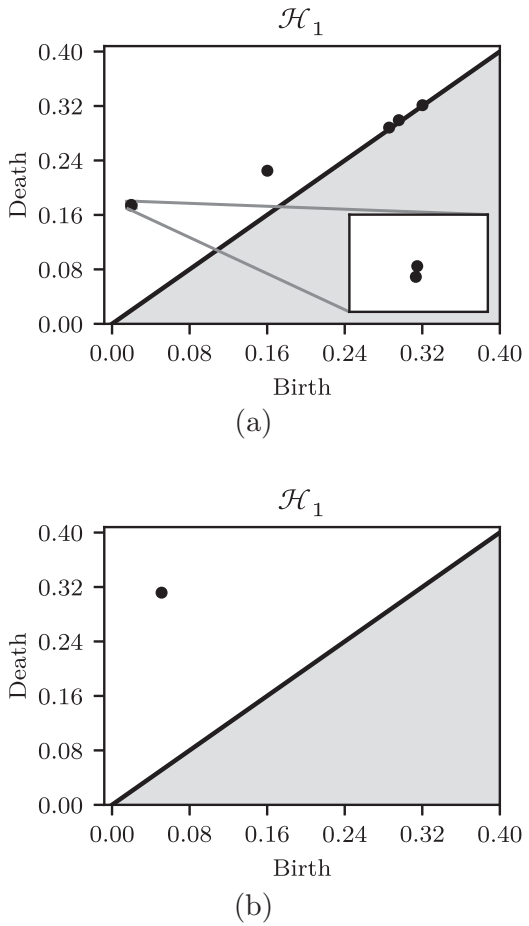

FIG. 12. Persistent homology of the observations in the experiment from Sec. IV E. (a) No correlation averaging. (b) Correlations averaged over ten shifts.

uniform, then we are in a setting where we believe our methods are applicable.

The question to ask is whether the persistent homology of the flag complex of the correlation-weighted graph of the observations closely approximates the persistent homology of the Rips complex of C.

In the above setup, define $g_k : M \times M \to \mathbb{R}$ by

$$g_k(y,z) = \int_M f(d(y,x))f(d(z,x))d\rho_k(x)$$

with ρ_k the probability density function for the random walk at time step k (assuming enough structure on M that the integral makes sense). The function g_k arises naturally as the probability

$$P[S_i(k) = 1 \cap S_j(k) = 1] = g_k(p_i, p_j)$$

and is the essential part of the estimated Pearson correlation $\mathrm{corr}(s_i,s_j)$. In the case that $X(k)$ is fully uniform, i.e., $d\rho_k = dx$, proving correctness of our recovered persistent homology amounts to showing the existence of a monotonically decreasing $h : \mathbb{R}^+ \to \mathbb{R}$ that makes

$$\begin{array}{ccc} M \times M & \xrightarrow{\ g\ } & \mathbb{R} \\ {\scriptstyle d}\downarrow & \nearrow_{h} & \\ \mathbb{R}^+ & & \end{array}$$

commute. In the Euclidean situation ($M = \mathbb{R}^n$ and dropping the assumptions on M) this is an easy calculus exercise, but the details need to be worked out for the more general situations.

A weakness of our method arises in cases where the system's position in state space is biased. The observed correlations then reflect this bias instead of the actual overlap of fields of elevated activity, unless one employs a model that can correct for the observational bias. In the neuroscience setting, such a bias can come from the animal tending to prefer certain regions of space or certain head angles.

Work should be done to give rigorous bounds on the persistence modules based on the statistical properties of the Pearson correlations $\mathrm{corr}(s_i,s_j)$, especially in the settings when the random walk distribution is not yet truly uniform (which clearly is the case in many real applications, especially if the observation time is short).

A later work [32] addresses real neuronal data with a method similar to ours. We believe both methods promise to provide useful insight into the geometric structures encoded in data from neuronal and related systems.

ACKNOWLEDGMENTS

We thank Y. Roudi for fruitful discussions. G.S. would also like to thank G.-A. Fuglstad for very valuable discussions on matters of probability. The later stages of the work of G.S. (writing up the paper) was partially supported by the NCCR Synapsy of the Swiss National Science Foundation. The research of M.B.B. was partially supported by the DFG Collaborative Research Center SFB/TR 109 "Discretization in Geometry and Dynamics." Persistent homology computations were performed using Bauer *et al.*'s DIPHA [33] and Nanda's PERSEUS [34] (version 4 beta) software.

[1] D. H. Hubel and T. N. Wiesel, J. Physiol. **148**, 574 (1959).
[2] G. B. Keller, T. Bonhoeffer, and M. Hübener, Neuron **74**, 809 (2012).
[3] W. Truccolo, U. T. Eden, M. R. Fellows, J. P. Donoghue, and E. N. Brown, J. Neurophysiol. **93**, 1074 (2005).
[4] I. H. Stevenson and K. P. Kording, Nat. Neurosci. **14**, 139 (2011).
[5] H. Akaike, in *Selected Papers of Hirotugu Akaike* (Springer, New York, 1998), pp. 199–213.
[6] G. Schwarz, Ann. Stat. **6**, 461 (1978).
[7] R. Ghrist, Bull. Am. Math. Soc. **45**, 61 (2008).
[8] H. Edelsbrunner and J. Harer, *Computational Topology: An Introduction* (American Mathematical Society, Providence, Rhode Island, USA, 2010).
[9] H. Edelsbrunner, *A Short Course in Computational Geometry and Topology* (Springer, New York, 2014).
[10] S. Y. Oudot, *Persistence Theory: From Quiver Representations to Data Analysis*, Mathematical Surveys and Monographs Vol.

209 (American Mathematical Society, Providence, Rhode Island, USA, 2015).
[11] C. Curto, Bull. Am. Math. Soc. **54**, 63 (2017).
[12] C. Curto and V. Itskov, PLoS Comput. Biol. **4**, e1000205 (2008).
[13] Y. Dabaghian, F. Mémoli, L. Frank, and G. Carlsson, PLoS Comput. Biol. **8**, e1002581 (2012).
[14] M. Arai, V. Brandt, and Y. Dabaghian, PLoS Comput. Biol. **10**, e1003651 (2014).
[15] C. Giusti, E. Pastalkova, C. Curto, and V. Itskov, Proc. Natl. Acad. Sci. USA **112**, 13455 (2015).
[16] J. O'Keefe and J. Dostrovsky, Brain Res. **34**, 171 (1971).
[17] E. Hermansen, Master's thesis, Norwegian University of Science and Technology, 2017.
[18] Y. Roudi, J. Tyrcha, and J. Hertz, Phys. Rev. E **79**, 051915 (2009).
[19] S. S. Borysov, Y. Roudi, and A. V. Balatsky, Eur. Phys. J. B **88**, 321 (2015).
[20] D. Taylor, F. Klimm, H. A. Harrington, M. Kramár, K. Mischaikow, M. A. Porter, and P. J. Mucha, Nat. Commun. **6**, 7723 (2015).
[21] G. Spreemann, B. Dunn, M. B. Botnan, and N. A. Baas, arXiv:1510.06629.
[22] M. Mézard and J. Sakellariou, J. Stat. Mech.: Theor. Exp. (2011) L07001.
[23] R. Ghrist, *Elementary Applied Topology* (Createspace, Philadelphia, Pennsylvania, USA, 2014).
[24] A. Hatcher, *Algebraic Topology* (Cambridge University Press, Cambridge, England, 2002).
[25] G. Azumaya, Nagoya Math. J. **1**, 117 (1950).
[26] W. Crawley-Boevey, J. Alg. Appl. **14**, 1550066 (2015).
[27] M. Kahle, Contemp. Math. **620**, 201 (2014).
[28] C. Carstens and K. Horadam, Math. Probl. Eng. **2013**, 815035 (2013).
[29] G. Petri, M. Scolamiero, I. Donato, and F. Vaccarino, PloS One **8**, e66506 (2013).
[30] J. P. Cunningham and M. Y. Byron, Nat. Neurosci. **17**, 1500 (2014).
[31] N. Otter, M. A. Porter, U. Tillmann, P. Grindrod, and H. A. Harrington, EPJ Data Sci. **6**, 17 (2017).
[32] E. Rybakken, N. Baas, and B. Dunn, arXiv:1711.07205.
[33] U. Bauer, M. Kerber, and J. Reininghaus, in *Proceedings of the Meeting on Algorithm Engineering & Expermiments* (Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, USA, 2014) pp. 31–38.
[34] K. Mischaikow and V. Nanda, Discrete Comput. Geom. **50**, 330 (2013).