

Two's company, three (or more) is a simplex: Algebraic-topological tools for understanding higher-order structure in neural data

Chad Giusti^{1,2}, Robert Ghrist^{2,3} and Danielle S. Bassett^{1,3}

¹Department of Bioengineering, University of Pennsylvania, Philadelphia, PA 19104

²Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104

³Department of Electrical & Systems Engineering, University of Pennsylvania,
Philadelphia, PA 19104

December 21, 2015

Abstract

The language of graph theory, or network science, has proven to be an exceptional tool for addressing myriad problems in neuroscience. Yet, the use of networks is predicated on a critical simplifying assumption: that the quintessential unit of interest in a brain is a dyad – two nodes (neurons or brain regions) connected by an edge. While rarely mentioned, this fundamental assumption inherently limits the types of neural structure and function that graphs can be used to model. Here, we describe a generalization of graphs that overcomes these limitations, thereby offering a broad range of new possibilities in terms of modeling and measuring neural phenomena. Specifically, we explore the use of *simplicial complexes*, a theoretical notion developed in the field of mathematics known as algebraic topology, which is now becoming applicable to real data due to a rapidly growing computational toolset. We review the underlying mathematical formalism as well as the budding literature applying simplicial complexes to neural data, from electrophysiological recordings in animal models to hemodynamic fluctuations in humans. Based on the exceptional flexibility of the tools and recent ground-breaking insights into neural function, we posit that this framework has the potential to eclipse graph theory in unraveling the fundamental mysteries of cognition.

Keywords. networks, topology, simplicial complex, filtration

The recent development of novel imaging techniques and the acquisition of massive collections of neural data make finding new approaches to understanding neural structure a vital undertaking. Network science is rapidly becoming an ubiquitous tool for understanding the structure of complex neural systems. Encoding relationships between objects of interest using graphs (Figure 1A-1B, Figure 4A) enables the use of a bevy of well-developed tools for structural characterization as well as inference of dynamic behavior. Over the last decade, network models have demonstrated broad utility in uncovering fundamental architectural principles [1, 2] and their implications for cognition [3] and disease [4]. Their use has led to the development of novel diagnostic biomarkers [4] and conceptual cognitive frameworks [5] that illustrate a paradigm shift in systems, cognitive, and clinical neuroscience: namely, that brain function and alteration are inherently networked phenomena.

All graph-based models consist of a choice of *vertices*, which represent the objects of study, and a collection of *edges*, which encode the existence of a relationship between pairs of objects (Figure 1A-1B, Figure 4A). However, in many real systems, such *dyadic* relationships fail to accurately capture the rich nature of the system's organization; indeed, even when the underlying structure of a system is known to be dyadic, its function is often understood to be polyadic. In large-scale neuroimaging, for example, cognitive functions appear to be performed by a distributed set of brain regions [6] and their interactions [3]. At a smaller scale, the spatiotemporal patterns of interactions between a few neurons is thought to underlie basic information coding [7] and explain alterations in neural architecture that accompany development [8].

Drawing on techniques from the field of *algebraic topology*, we describe a mathematically well-studied generalization of graphs called *simplicial complexes* as an alternative, often preferred method for encoding non-dyadic relationships (Figure 4). Different types of complexes can be used to encode co-firing of neurons [9], co-activation of brain areas [10], and structural and functional connections between neurons or brain regions [11] (Figure 5). After choosing the complex of interest, quantitative and theoretical tools can be used to describe, compare, and explain the statistical properties of their structure in a manner analogous to graph statistics or network diagnostics.

We then turn our attention to a method of using additional data, such as temporal processes or frequency of observations, to decompose a simplicial complex into constituent pieces, called a *filtration* of the complex (Figure 1F-1H). Filtrations reveal more detailed structure in the complex, and provide tools for understanding how that structure arises (Figure 7). They can also be used as an alternative to thresholding a weighted complex, providing a principled approach to binarizing which retains all of the data in the original weighted complex.

In what follows, we avoid introducing technical details beyond those absolutely necessary, as they can be found elsewhere [12, 13, 14].

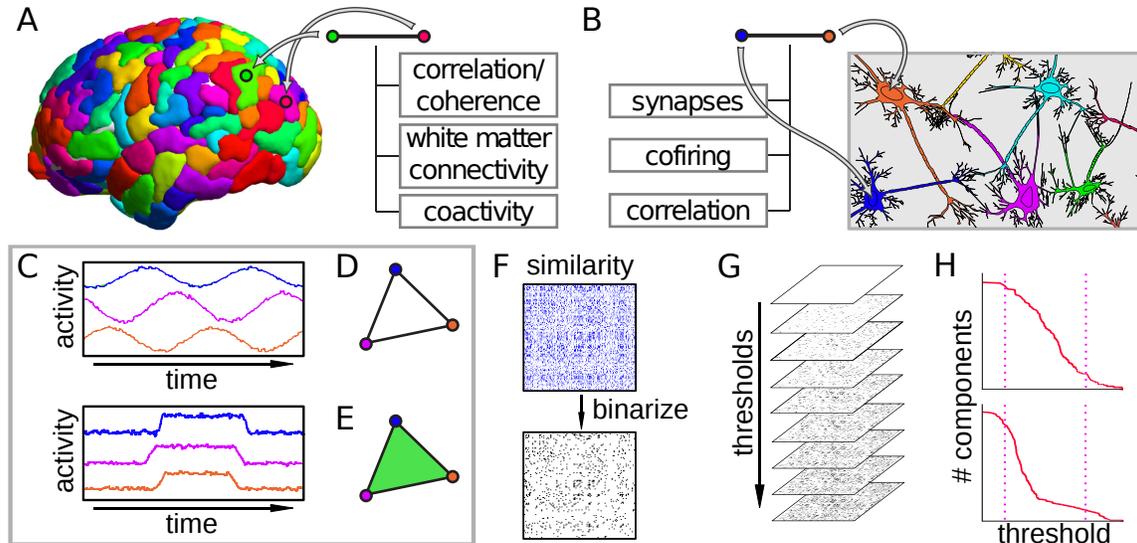


Figure 1: Extensions of network models provide insights into neural data. (A) Network models are increasingly common for the study of whole-brain activity. (B) Neuron-level networks have been a driving force in the adoption of network techniques in neuroscience. (C) Two potential activity traces for a trio of neural units. (top) Activity for a “pacemaker”-like circuit, whose elements are pairwise active in all combinations but never as a triple. (bottom) Activity for units driven by a common strong stimulus, thus are simultaneously coactive. (D) A network representation of the coactivity patterns for either population in (C). Networks are capable of encoding only dyadic relationships, so do not capture the difference between these two populations. (E) A *simplicial complex* model is capable of encoding higher order interactions, thus distinguishing between the top and bottom panels in (C). (F) A similarity measure for elements in a large neural population is encoded as a matrix, thought of as the adjacency matrix for a complete, weighted network, and binarized using some threshold to simplify quantitative analysis of the system. In the absence of complete understanding of a system, it is difficult or impossible to make a principled choice of threshold value. (G) A *filtration* of networks is obtained by thresholding at every possible entry and arranging the resulting family of networks along an axis at their threshold values. This structure discards no information from the original weighted network. (G) Graphs of the number of connected components as a function of threshold value for two networks reveals differences in their structure: (top) homogeneous network *versus* (bottom) a modular network. (dotted lines) Thresholding near these values would suggest inaccurately that these two networks have similar structure.

Motivating examples

Before we dive into describing the tools and how they have been used, we begin with a pair of simple thought experiments which highlight more explicitly the reasons we consider these techniques to be valuable for the study of neural systems.

First, imagine a simple neural system consisting of three brain regions (or neurons) with unknown connectivity. One possible activity profile for such a population includes some sort of sequential information processing loop or “pacemaker” like circuit, where the regions activate in a rotating order (Figure 1C, top). A second is for all three of the regions to be active simultaneously when engaged in certain computations, and otherwise quiescent or uncorrelated (Figure 1C, bottom). In either case, an observer would find the

activity of all three possible pairs of regions to be strongly correlated. Because a network can only describe dyadic relationships between population elements, any binary coactivity network constructed from such observations would necessarily be identical for both (Figure 1D). However, a more versatile language could distinguish the two by explicitly encoding the triple coactivity pattern in the second example (Figure 1E). The framework of *simplicial complexes* (Figure 4B-4D) is such a language, a straightforward extension of the formalism of graph theory that allows one to describe relations between arbitrarily large sub-populations without sacrificing computability or access to many of the fundamental tools of network science. Further, the richer structure inherent in simplicial complexes has driven the development of correspondingly more powerful mathematical techniques for detecting and analyzing the structure of the systems they encode. These methods provide a quantitative architecture through which to address modern questions about complex and emergent behavior in neural systems.

Second, consider a much larger neural system, consisting of several hundred units, whose activity is summarized as a correlation or coherence matrix (Figure 1F, top). It is common practice to binarize such a matrix by thresholding it at some value, taking entries above that value to be “significant” connections, and to study the resulting, much sparser network (Figure 1F, bottom). Selecting this significance level is problematic, particularly when the underlying system is not thoroughly understood and low-impact effects that might be dismissed as noise are potentially important to its function. One method for working around this difficulty is to take several thresholds and study the results separately. However, this approach still discards most of the information contained in the edge weights, much of which can be of inherent value in understanding the system. We propose instead the use of *filtrations*, which record the results of every possible binarization of the network, along with the associated threshold value (Figure 1G). Filtrations not only retain all of the information in the original weighted networks, but unfold that information into a more accessible form, allowing one to lift any measure of structure in networks (or simplicial complexes) to “second order” measures as functions of edge weight (Figure 1H). Such functions carry information, for example, in their rate of change, where sudden phase transitions in network structure as one varies the threshold can indicate the presence of modules or rich clubs in networks (Figure 1H). Alternately, the area under such curves was used in [15] to detect geometric structure in the activity of hippocampal neural populations (Figure 3). Further, even more delicate information can be extracted from the filtration by tracking the *persistence* of individual structures in the graphs (such as components) as the threshold varies (Figure 7C).

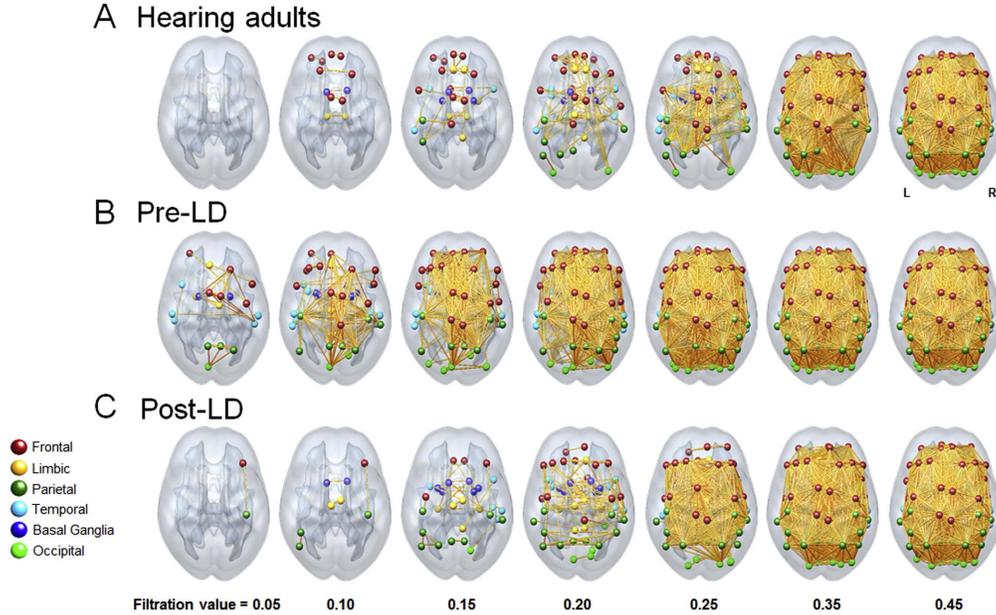


Figure 2: Filtered brain networks constructed from interregional correlations of density from MRI detect differences in hearing and deaf populations. Density correlation networks obtained from (A) hearing, (B) prelingual deaf, and (C) postlingual deaf adults. Differences in the evolution of network components across groups as the threshold parameter varies provides insight into differences in structure. It is unclear how one would select a particular threshold which readily reveals these differences without *a priori* knowledge of their presence. Figure reproduced with permission from [16].

A Growing Literature

Before we begin a careful discussion of the mathematical concepts described above, we provide an overview of the existing literature, which can be roughly divided into two branches:

Building simplicial complexes to describe neural coding and network properties. In [9], a novel kind of simplicial complex derived from neural data was introduced to show how hippocampal place cell activity can, in principle, be used to reconstruct the topology (and much of the geometry) of the represented environment. Further work in this direction has explored how such a mechanism would interact with known dynamics in the hippocampus [17], how spatial maps would form through such a mechanism [18], and how to apply such techniques to observed cell activity to understand place cell function [19] or decode probabilistic maps of the environment [20]. In [21], a filtration of simplicial complexes was used to study complex interregional patterns in fMRI recordings, and in [23], similar methods were used to differentiate *in vivo* cortical cell cultures into functional sub-networks under various system conditions. Finally, a version of simplicial complexes that includes a notion of *directedness* has been used to investigate the relationship between simulated structural and functional neural networks (Dlotko et al., unpublished).

Using measurements of filtrations to characterize brain architecture or state. One of the earliest

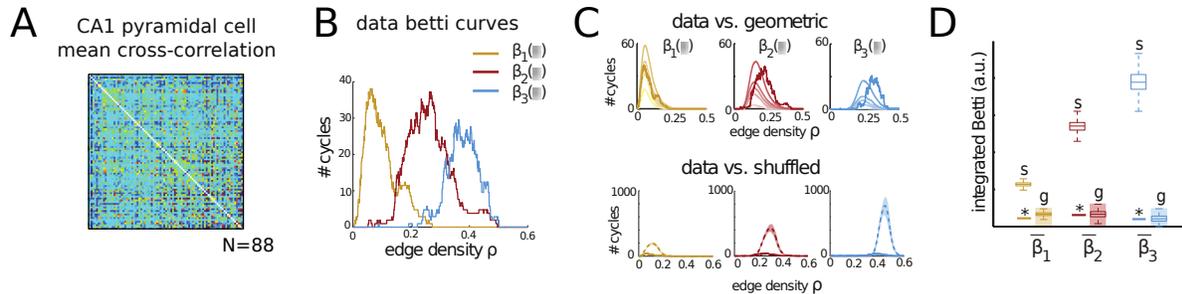


Figure 3: Betti numbers detect the existence of geometric organizing principles in neural population activity from rat hippocampus. (A) Mean cross correlation of N=88 rat CA1 pyramidal cells activity during spatial navigation. (B) Betti numbers as a function of graph edge density ($\#$ edges / possible $\#$ edges) for the clique complex of the pairwise correlation network in (A). (C) Comparison of data Betti numbers (thick lines) to model random networks with (top) geometric weights given by decreasing distance between random points in Euclidean space and (bottom) with no intrinsic structure obtained by shuffling the entries of the correlation matrix. (D) Integrals of the curves from panel B show that the data (thick bars) lie in the geometric regime (g) and that the unstructured network model (s) is fundamentally different ($p < 0.001$). Similar geometric organization was observed in non-spatial behaviors such as REM sleep. Figure reproduced with permission from [15].

applications of algebraic topology to neural data was to the study of activity in the macaque primary visual cortex [24], where different distributions of algebraic-topological features were observed across recordings of spontaneous activity and during exposure to natural images. Later work applied these methods to imaging data, using cortical thickness to differentiate ASD subjects from controls [25], to understand the spatial structure of afferent neuron terminals in crickets [26], and to detect human age and gender from the structure of brain artery trees [27]. Also common has been the use of correlation of observed neuronal population activity to construct weighted graphs, from these to construct filtered simplicial complexes and then compute algebraic-topological measurements to be used as a discriminator of classes of subjects. Focusing on persistence of components, this technique was used in [28] to classify pediatric ADHD, ASD and control subjects; in [29] to differentiate mouse models of depression from controls; in [30] to differentiate epileptic rat models from controls; and in [16] to study morphological correlations in adults with hearing loss (Figure 2). Similar work with persistence of more intricate algebraic-topological features computed from fMRI recordings distinguishes subjects under psilocybin condition from controls [31]. More recently, these techniques have been adapted to detect structure in the information encoded by a population through observations of its activity [15] (Figure 3).

The field of topological neuroscience is both very new and very small, yet it already offers an array of powerful new quantitative approaches for addressing the unique challenges inherent in understanding neural systems and it has begun making substantial contributions. In recent years, there have been a number of

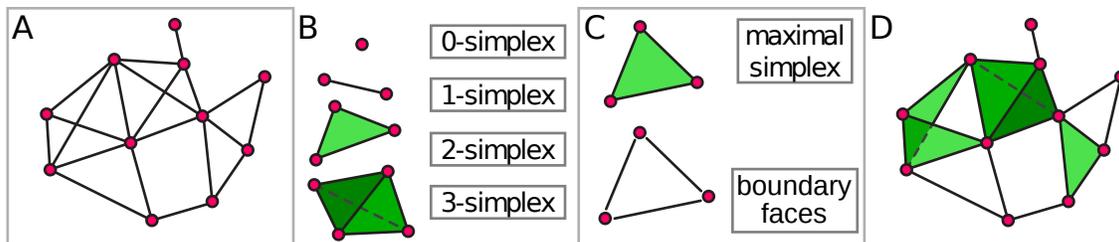


Figure 4: Simplicial complexes generalize network models. (A) A graph encodes elements of a neural system as vertices and dyadic relations between them as edges. (B-C) Simplicial complex terminology. A simplicial complex is made up of *vertices* and *simplices*, which are defined in terms of collections of vertices. (B) A k -simplex can be thought of as the convex hull of $(k + 1)$ vertices. (C) The *boundary* of a simplex consists of all possible subsets of its constituent vertices, called its *faces*, which are themselves required to be simplices in the complex. A simplex which is not in the boundary of any other simplex is called *maximal*. (D) A simplicial complex encodes polyadic relations through its simplices. Here, in addition to the dyadic relations specified by the edges, the complex specifies one four-vertex relation and three three-vertex relations. The omission of larger simplices where all dyadic relations are present, such as the three bottom-left vertices or the four top-left vertices, encodes structure that cannot be specified using network models.

innovative collaborations between mathematicians interested in applying topological methods and researchers in a variety of biological disciplines, including the discovery of new genetic markers for breast cancer survival [32], measurement of structure and stability of biomolecules [33, 34], new frameworks for understanding viral evolution [35], and characterization of dynamics in gene regulatory networks [36]. This wide-spread interest is an untapped resource for empirical neuroscientists which promises to facilitate both direct applications of existing techniques and the collaborative construction of novel tools specific to their needs.

We devote the remainder of the paper to a careful exposition of these techniques, highlighting specific ways that they may or have already been used to address questions of interest to neuroscientists.

Mathematical framework: Simplicial complexes

We begin with a short tutorial on simplicial complexes, and illustrate the similarities and differences with graphs.

A *simplicial complex*, like a graph, consists of a set of vertices and a specified collection of subsets of those vertices, called *simplices*, subject to the mild restriction that any subset of a simplex must also be a simplex. Observe that any graph is automatically a simplicial complex with all simplices being either vertices or pairs (edges). General simplicial complexes possess more subtle information.

Just as one can represent a graph as a collection of points and line segments between them, one can represent the simplices in a simplicial complex as a collection of solid regions connecting vertices (Figure 4D). Under this geometric interpretation, a single vertex is a zero-dimensional point, while two distinct points

define a one-dimensional line segment, three points a two-dimensional triangle, and so on. Terminology for simplices is derived from this geometric representation: a simplex on $(n + 1)$ vertices is called an n -simplex and is viewed as spanning an n -dimensional region. Further, the requisite subsets of a simplex represent regions in the geometric boundary of the simplex (Figure 4E), so these subsets of a simplex are called its *faces*.

Because any given simplex is required to “contain all of its faces”, to identify a complex it is sufficient to specify only the *maximal simplices*, those which do not appear as faces of another simplex (Figure 4E). This dramatically reduces the amount of data necessary for working with simplicial complexes, which helps make computations feasible.

In real-world systems, simplicial complexes possess richly structured patterns that can be detected and characterized using recently developed computational tools from algebraic topology [37, 38], just as graph theoretic tools can be used to study networks. Importantly, these tools reveal much deeper properties of the relationships between vertices than graphs, and many are constructed not only to see structure in individual simplicial complexes, but also to help one understand how two or more simplicial complexes compare or relate to one another. These capabilities naturally enable the study of complex dynamic structure in neural systems, and formalize statistical inference via comparisons to null models.

Formal Definitions

An (*abstract*) *simplicial complex* X is a pair of sets: V_X , called the *vertices*; and S_X , called the *simplices*, each of which is a finite subset of V_X , subject to the requirement that if σ is in S_X , then every subset τ of σ is also in S_X . A simplex with n elements is called an $(n - 1)$ -*simplex*, and subsets $\tau \subset \sigma$ are *faces* of σ .

How do we encode neural data?

To demonstrate the broad utility of this framework, we turn to describing a selection of the many types of simplicial complexes that can be constructed from data: the clique complex, the concurrence complex [21, 9, 39], its Dowker dual [39], and the independence complex [14]. In each case, we describe the relative utility in representing different types of neural data – from spike trains measured from individual neurons to BOLD activations measured from large-scale brain areas.

Clique Complex. One straightforward method for constructing simplicial complexes begins with a graph where vertices represent neural units and edges represent structural or functional connectivity between those units (Figure 4A-4B). Next, one replaces every *clique* (all-to-all connected subgraph) by a simplex on the vertices participating in the clique (Figure 5A). This procedure produces a *clique complex*, which encodes the same information as the underlying graph, but additionally completes the skeletal network to

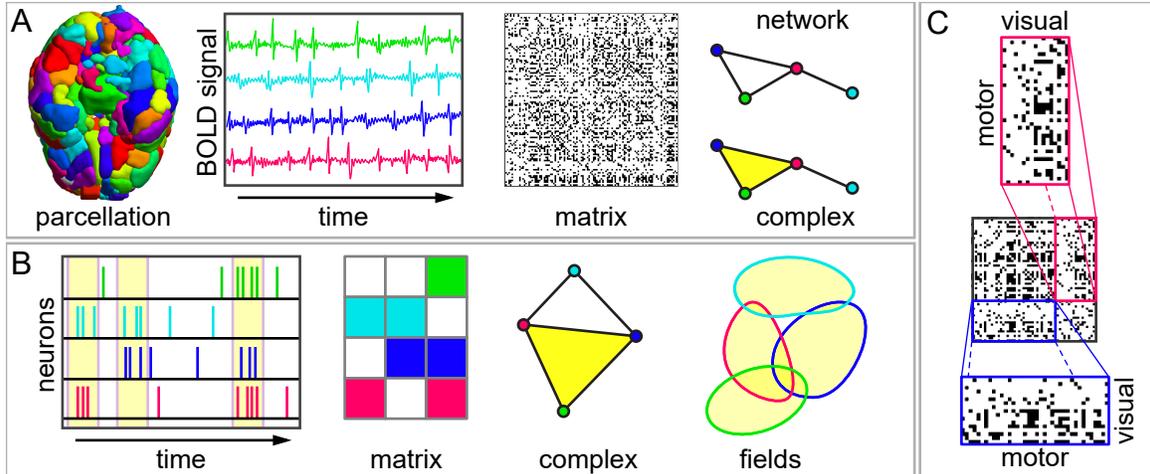


Figure 5: Simplicial complexes encode diverse neural data modalities. (A) Correlation or coherence matrices between regional BOLD time series can be encoded as a type of simplicial complex called a *clique complex*, formed by taking every complete (all-to-all) subgraph in a binarized functional connectivity matrix to be a simplex. (B) Coactivity patterns in neural recordings can be encoded as a type of simplicial complex called a *concurrence complex*. Here, we study a binary matrix in which each row corresponds to a neuron and each column corresponds to a collection of neurons that is observed to be coactive at the same time (yellow boxes) – i.e., a simplex. (C) Thresholded coherence between the activity patterns of motor regions and visual regions in human fMRI data during performance of a motor-visual task [44]. (top) We can construct a concurrence complex whose vertices are motor regions and whose simplices are families of motor regions whose activity is strongly coherent with a given visual region. (bottom) We can also construct a *dual* complex whose vertices are families of motor regions. The relationship between these two complexes carries a great deal of information about the system [39].

the fullest simplicial structure. The utility of this structure was recently demonstrated in the context of neural activity measured in rat hippocampal pyramidal cells during both spatial and non-spatial behavior (including REM sleep) [15] (Figure 3). In contrast to graph statistics, the pattern of simplices revealed the presence of geometric structure in only the information encoded in neural population activity correlations that – surprisingly – could be identified and characterized independently from the animal’s position. This application demonstrates that simplicial complexes are sensitive to organizational principles that are hidden to graph statistics, and can be used to infer parsimonious rules for information encoding in neural systems.

Simplicial Complex Type	Utility
Graph	General framework for encoding dyadic relations
Clique Complex	Canonical polyadic extension of existing network models
Concurrence Complex/Dual	Relationships between two variables of interest e.g., time and activity, or activity in two separate regions
Independence Complex	Structure where non-membership satisfies the simplex property e.g., communities in a network

Table 1: Comparison of sample types of simplicial complexes for encoding neural data.

Clique complexes precisely encode the topological features present in a graph. However, other types of simplicial complexes can be used to represent information that *cannot* be so encoded in a graph.

Concurrence Complex. Using cofiring, coactivity, or connectivity as before, let us consider relationships between two different sets of variables. For example, we can consider (i) neurons and (ii) times, where the relationship is given by a neuron firing in a given time (Figure 5B) [9]; a similar framing exists for (i) brain regions and (ii) times, where the relationship is given by a brain region being active at a given time [21]. Alternatively, we can consider (i) brain regions in the motor system and (ii) brain regions in the visual system, where the relationship is given by a motor region displaying similar BOLD activity to a visual region (Figure 5C) [22]. In each case, we can record the patterns of relationships between the two sets of variables as a binary matrix, where the rows represent elements in one of the variables (e.g., neurons) and the columns the other (e.g., times), with non-zero entries corresponding to the row-elements in each column sharing a relation (e.g., firing together at a single time). The *concurrence complex* is formed by taking the rows of such a matrix as vertices and the columns to represent maximal simplices consisting of those vertices with non-zero entries [39]. A particularly interesting feature of this complex is that it remains naive to coactivity patterns that do not appear, and this naivety plays an important role in its representational ability; for example, such a complex can be used to decode the geometry of an animal’s environment from observed hippocampal cell activity [9].

Moving to simplicial complex models provides a dramatically more flexible framework for specifying data encoding than simply generalizing graph techniques. Here we describe two related simplicial complex constructions from neural data which cannot be represented using network models.

Dowker Dual. Beginning with observations of coactivity, connection or cofiring as before, one can choose to represent neural units as simplices whose constituent vertices represent patterns of coactivity in which the unit participates. Expressing such a structure as a network would necessitate every neural unit participating in precisely two activity patterns, an unrealistic requirement, but is straightforward in the simplicial complex formalism. Mathematically speaking, one can think of the matrix encoding this complex as the transpose of the matrix encoding the concurrence complex; such “dual” complexes are deeply related to one another, as first observed by Dowker [39]. Critically, this formulation refocuses attention (and the output of various vertex-based statistical measures) from individual neural units to patterns of coactivity.

Independence Complex. It is sometimes the case that an observed structure does not satisfy the simplicial complex property, but its complement does. One example of interest is the collection of *communities* in a network [40, 41]: communities are subgraphs of a network whose vertices are more densely connected to

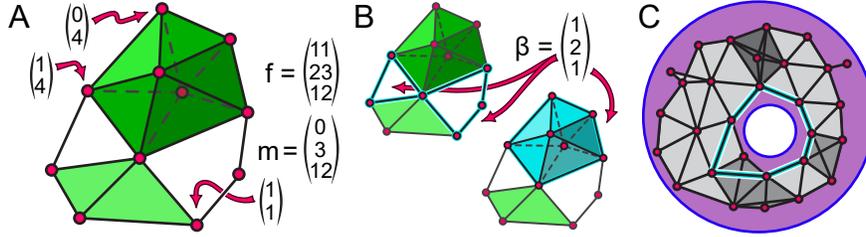


Figure 6: Quantifying the structure of a simplicial complex. (A) Generalizations of the degree sequence for a simplicial complex. Each vertex has a *degree* vector giving the number of maximal simplices of each degree to which it is incident. The *f-vector* gives a list of how many simplices of each degree are in the complex, and the *maximal simplex distribution* records only the number of maximal simplices of each dimension. (B) Closed cycles of dimension 1 and 2 in the complex from panel A. (left) There are two independent 1-cycles (cyan) up to deformation through 2-simplices, and (right) a single 2-cycle (cyan) enclosing a 3-d volume. The Betti number vector β gives an enumeration of the number of n -cycles in the complex, here with $n = 0, 1$ and 2 ; the single 0-cycle corresponds to the single connected component of the complex. (C) Schematic representation of the reconstruction of the presence of an obstacle in an environment using a concurrence complex constructed from place cell cofiring [9]. By choosing an appropriate cofiring threshold, based on approximate radii of place cell receptive fields, there is a single 1-cycle (cyan), up to deformation through higher simplices, indicating a large gap in the receptive field coverage where the obstacle appears.

one another than expected in an appropriate null model. The collection of vertices in the community is not a simplex, because removing densely connected vertices can cause the community to dissolve. Thus, community structure is well-represented as a *hypergraph* [42], though such structures are much less natural and useful than simplicial complexes. However, in this setting simplices can be taken to be all vertices *not* in a community. Such a simplicial complex is again essentially a concurrence complex: simply negate the binary matrix whose rows are elements of the network and columns correspond to community membership. Such a complex is called an *independence complex* [14], and can be used to study properties of a system's community structure such as dynamic flexibility [43, 44].

Together, these different types of complexes can be used to encode a wide variety of relationships (or lack thereof) among neural units or coactivity properties in a simple matrix that can be subsequently interrogated mathematically. This is by no means an exhaustive list of complexes of potential interest to the neuroscience community; for further examples, we recommend [12, 14].

How do we measure the structure of simplicial complexes?

Just as with network models, once we have effectively encoded neural data in a simplicial complex, it is necessary to find useful quantitative measurements of the resulting structure to draw conclusions about the neural system of interest. Because simplicial complexes generalize graphs, many familiar graph statistics can be extended in interesting ways to simplicial complexes. However, algebraic topology also offers a host

of novel and very powerful tools that are native to the class of simplicial complexes, and cannot be derived from well known graph theoretical constructs.

Graph Theoretical Extensions. First, let us consider how we can generalize familiar graph statistics to the world of simplicial complexes. The simplest local measure of structure – the *degree* of a vertex – naturally becomes a vector-measurement whose entries are the number of maximal simplices of each size in which the vertex participates (Dlotko et al., unpublished) (Figure 6A). Although a direct extension of the degree, this vector is perhaps more intuitively thought of as a generalization of the *clustering coefficient* of the vertex: in this setting we can distinguish empty triangles, which represent three dyadic relations but no triple-relations, from 2-simplices which represent clusters of three vertices (and similarly for larger simplices).

Just as we can generalize the degree, we can also generalize the degree distribution. Here, the *simplex distribution* or *f-vector* is the global count of simplices by size, which provides a global picture of how tightly connected the vertices are; the *maximal simplex distribution* collects the same data for maximal faces (Figure 6A). While these two measurements are related, their difference occurs in the complex patterns of overlap between simplices and so together they contain a great deal of structural information about the simplicial complex. Other local and global statistics such as *efficiency* and *path length* can be generalized by considering paths through simplices of some fixed size, which provides a notion of *robust connectivity* between vertices of the system (Dlotko et al., unpublished); alternately, a path through general simplices can be assigned a strength coefficient depending on the size of the maximal simplices through which it passes.

Algebraic-Topological Methods. However, while these generalizations are possible, and likely of significant interest to the neuroscience community, they are not the fundamental statistics originally developed to characterize simplicial complexes. In their original context, simplicial complexes were used to study shapes, using *algebraic topology* to measure global structure.

The most commonly used and computationally tractable of these measurements is the (*simplicial*) *homology* of the complex¹, which is actually a sequence of measurements. The n^{th} *homology* of a simplicial complex is the collection of (*closed*) *n-cycles* formed from *n-simplices* (Figure 6B), up to a notion of equivalence². Such cycles can be thought of as characterizing “holes” in various dimensions³, and are an example of global structure arising from local structure; simplices arrayed across multiple vertices must coalesce in a particular fashion to encircle a hole not filled in by other simplices. In many settings, a powerful summary statistic is simply a count of the number of inequivalent cycles of each dimension appearing in the complex. These counts are called *Betti numbers*, and collect them as a vector β .

¹Names of topological objects have a seemingly pathological tendency to conflict with terms in biology, so long have the two subjects been separated. Mathematical homology has no *a priori* relationship to the usual biological notion of homology.

²Two *n-cycles* are equivalent if they differ by the boundary of some collection of $(n + 1)$ -simplices.

³The actual definition of a cycle is more subtle and requires careful discussion. We refer the interested reader to the aforementioned expositions [12, 13].

In the context of neural data, the presence of multiple homology classes indicates potentially interesting structure whose interpretation depends on the meaning of the vertices and simplices in the complex. For example, the open triangle in the complex of Figure 5B is a 1-cycle representing pairwise coactivity of all of the constituent neurons but a lack of triple coactivity; thus, the reconstructed receptive field model includes no corresponding triple intersection, indicating a hole or obstacle in the environment. In the context of regional coactivity in fMRI, such a 1-cycle might correspond to observation of a distributed computation that does not involve a central hub. Cycles of higher dimension are more intricate constructions, and their presence or absence can be used to detect a variety of other more complex, higher-order features.

Formal Definitions

In order to compute with simplicial complexes, we convert their assembly implicit assembly instructions into linear algebra as follows. For a simplicial complex X , define a sequence of vector spaces $C_n(X)$ (over the finite field $\mathbb{F}_2 = \{0, 1\}$) with bases the n -simplices of X , writing $[\sigma]$ for the basis element corresponding to the simplex σ . Then define the *boundary* maps as linear transformations $\partial_n: C_n(X) \rightarrow C_{n-1}(X)$ assigning to each n -simplex $[\sigma]$ the formal sum $\sum_{[\tau]} [\tau]$ over all $(n-1)$ -simplex faces $[\tau]$ of σ . The *n th homology* of X is defined as the quotient vector space $H_n(X) = \ker \partial_n / \text{im } \partial_{n+1}$, and the *n th Betti number* is its dimension $\beta_n = \dim H_n(X)$.

Additional Tools to Assess Hierarchical and Temporal Structure

In previous sections we have seen how we can construct simplicial complexes from neural data and interrogate the structure in these complexes using both extensions of common graph theoretical notions and completely novel tools drawn from algebraic topology. We close the mathematical portion of this exposition by discussing a computational process that is common in algebraic topology and that directly addresses two critical needs in the neuroscience community: (i) the assessment of hierarchical structure in relational data via a principled thresholding approach, and (ii) the assessment of temporal properties of stimulation, neurodegenerative disease, and information transmission.

Filtrations to Assess Hierarchical Structure in Weighted Networks. One of the most common features of network data is a notion of *strength* or *weight* of connections between nodes. In some situations, like measurements of correlation or coherence of activity, the resulting network has edges between every pair of nodes and it is common to *threshold* the network to obtain some sparser, unweighted network whose edges correspond to “significant” connections [45]. However it is difficult to make a principled choice of threshold [46, 47, 48, 49, 50, 51], and the resulting network discards a great deal of information. Even in the case of sparse weighted networks, many metrics of structure are defined only for the underlying unweighted network,

so in order to apply the metric, the weights are discarded and this information is again lost [52].

Here, we describe a technique that is commonly applied in the study of weighted simplicial complexes which does not discard any information. Generalizing weighted graphs, *weighted simplicial complexes* have assigned to each of simplex (including vertices) a numeric *weight*, subject to the restriction that the weight of a simplex is no larger than that of any of its faces. That is, an observed relationship between any subset of a population is at least as strong as that observed among any larger subpopulation containing it. Given a weighted simplicial complex, a *filtration* of complexes can be constructed by consecutively applying each of the weights as thresholds and labeling each complex by the weight at which it was binarized. The resulting sequence of complexes retains all of the information in the original weighted complex, but one can apply metrics that are undefined or difficult to compute for weighted complexes to the entire collection, thinking of the resulting values as a function parameterized by the weights of the original complex (Figure 7D). However, it is also the case that these unweighted complexes are related to one another, and more sophisticated measurements of structure, like homology, can

exploit these relations to extract much finer detail of the evolution of the complexes as the threshold varies (Figure 7C). We note that the omni-thresholding approach utilized in constructing a filtration is a common theme among other recently developed methods for network characterization, including cost integration [46] and functional data analysis [47, 21].

The formalism described above provides a principled framework to translate a weighted graph or weighted simplicial complex into a family of unweighted graphs or complexes that retain all information in the weighting by virtue of their relationships to one another. However, filtrations are much more generally useful: for example, they can be used to assess the dynamics of neural processes.

Filtrations to Assess Temporal Dynamics of Neural Processes in Health and Disease

Many of the challenges faced by cutting edge experimental techniques in the field of neuroscience are driven by the underlying difficulties implicit in assessing temporal changes in complex patterns of relation-

Formal Definitions

A *filtration* of a simplicial complex X is a sequence of simplicial subcomplexes of the form

$$\emptyset = X_0 \subset X_1 \subset X_2 \subset \dots \subset X_M = X,$$

where \subset denotes inclusion as a subcomplex.

A simplicial complex is *weighted* if each simplex σ is assigned a real-valued weight $w(\sigma)$ so that if τ is a face of σ , then $w(\tau) \geq w(\sigma)$. Every weighted simplicial complex can be naturally converted into a filtration. Because the complex contains a finite number of simplices, the weight function w takes on only finitely many values, $w_1 \geq w_2 \geq \dots \geq w_M$. We can thus construct a subcomplex X_i of X by considering the subcomplexes of X for which $w(\sigma) \geq w_i$. These subcomplexes fit together to form a filtration (Figure 7B).

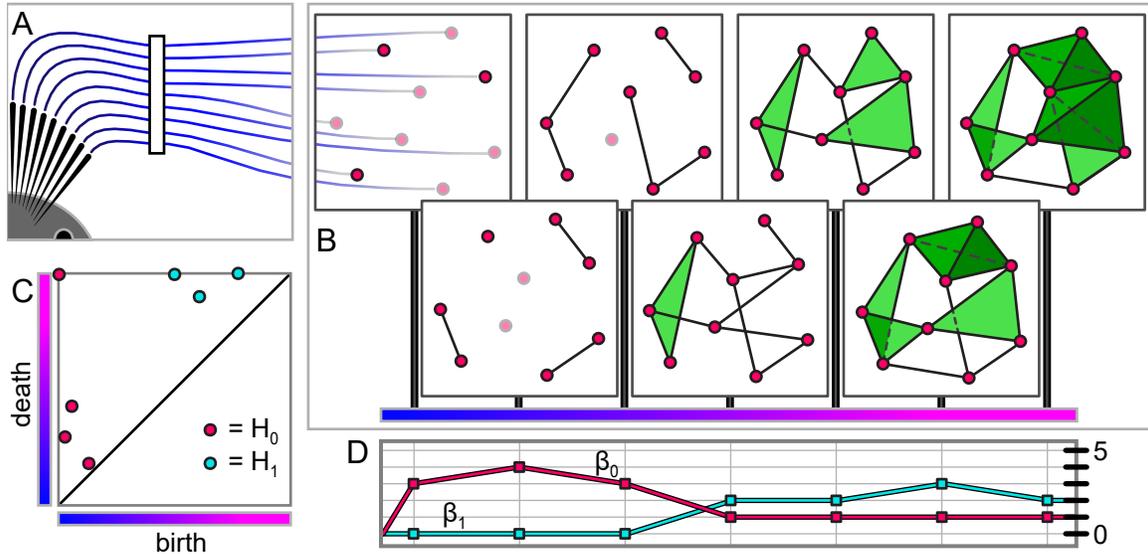


Figure 7: Filtrations of a weighted simplicial complex measure dynamic network properties. (A) A neural system can be stimulated in precise locations using electrical, magnetic or optogenetic methods and the resulting activity recorded. (B) A filtration of simplicial complexes is built by recording as maximal faces all patterns of coactivity observed up to a given time. Filtrations can be constructed from any weighted simplicial complex by thresholding using some *filtration parameter*. (C) A *persistence diagram* recording the appearance (“birth”) and disappearance or merging (“death”) of homology cycles throughout the filtration in panel B. Cycles on the top edge of the diagram are those that do not die. Tracking equivalent cycles through the filtration provides information about the evolution of structure as the filtration parameter changes. (D) *Betti curves* are the Betti numbers for each complex in the filtration of panel B represented as functions of time. Such curves can be constructed for any numerical measurement of the individual unweighted simplicial complexes in the filtration and provide a more complete description of structure than the individual measurements taken separately.

ships. For example, with new optogenetics capabilities, we can stimulate single neurons or specific groups of neurons to control their function [53]. Similarly, advanced neurotechnologies including microstimulation, transcranial magnetic stimulation, and neurofeedback enable effective control over larger swaths of cortex [54, 55]. With the advent of these technologies, it becomes imperative to develop computational tools to quantitatively characterize and assess the impact of stimulation on system function, and more broadly, to understand how the structure of a simplicial complex affects the transmission of information.

To meet this need, one can construct a different type of filtration, such as that introduced in [56] in the context of graphs: construct a sequence of simplicial complexes with a time parameter, labeling each simplex as “on” or “off” at each time, and require that once simplices “turn on” they remain so indefinitely. If the function has the further requirement that in order for a simplex to be active, all of its faces must be as well, then a filtration is obtained by taking all active simplices at each time. Such functions are quite natural to apply to the study of the pattern of neurons or neural units that are activated following stimulation.

Interestingly, this type of filtration is also a natural way in which to probe and reason about models of neurodegenerative disease such as the recently posited *diffusion model* of fronto-temporal dementia [57, 58]. Here, critical network epicenters form points of vulnerability that are effected early in the disease, and from which toxic protein species travel via a process of transneuronal spread. Indeed, these filtrations were first introduced in the context of contagion models [56], where a simplex becomes active once sufficiently many nearby simplices are active.

Measuring the Structure of Filtrations

Assuming we have encoded our data in an appropriate filtration, guided by our scientific hypothesis of interest, we might next wish to quantitatively characterize and measure the structure in those filtrations. It is important to note that any given measure of the structure of a simplicial complex can be applied to each complex in a filtration in turn, producing a function from the set of weights appearing in the complex to the set of values the measure can take (Figure 7D). This function is a new measure of the structure of the complex which does not rely on thresholds and can highlight interesting details that would not be apparent at any fixed threshold (or small range of thresholds), as well as being more robust to perturbations in the weights than measurements of any individual complex in the filtration.

Of particular interest in this setting are those quantitative measures whose evolution can be explicitly understood in terms of the underlying map of complexes, as then we can exploit the sequence of maps in the filtration to gain a more refined picture of the structure present in the weighted complex. Central among these in terms of current breadth of application and computability is *persistent homology*, which extends the homology of individual complexes to filtrations by tracking how equivalent cycles evolve through the growing filtration. Increasing the parameter sends each cycle to some cycles in the next simplicial complex

in the filtration. Therefore, the sequence of subcomplexes in the filtration is transformed by homology into an inter-related family of evolving cycles. Inside this sequence, cycles have well-defined *birth* and *death* weights, between which they evolve as new simplices are added, changing their form. This information is often encoded in *persistence diagrams* for each degree n (Figure 7C), which give a schematic overview of where the cycles are born and die. Understanding these *persistence lifetimes* of cycles can provide critical information about how the system is arranged. These techniques have been applied to uncover structure in the space of natural images [59], to detect subject gender from the shape of brain artery trees [27], and to identify statistically anomalous multi-region activation patterns in fMRI recordings [21].

Conclusion

We sit at a unique juncture in time, in which it is critical to support the principled development of novel computational tools that are not merely modular, but instead are tuned to address specific neuroscientific challenges at hand. With the feverish rise of data being collected from neural systems across species and spatial scales, mathematicians and experimental scientists must necessarily engage in deeper conversation about how meaning can be drawn from minutia. Such conversations will inevitably turn to the common understanding in the neurosciences that it is not necessarily the individual objects of study themselves, but their relations to one another, that provide the real structure of human and animal thought. Though originally developed for entirely different purposes, the algebraic topology of simplicial complexes provides a quantitative methodology uniquely suited to address these needs.

References

1. Bassett, D. S. and Bullmore, E. T. (2006). Small-world brain networks. *Neuroscientist* 12, 512–523.
2. Bullmore, E. T. and Bassett, D. S. (2011). Brain graphs: graphical models of the human brain connectome. *Ann. Rev. Clin. Psych.* 7, 113–140.
3. Medaglia, J. D. *et-al.* (2015). Cognitive Network Neuroscience. *J. Cogn. Neurosci.* 27, 1471–1491.
4. Stam, C. J. (2014). Modern network science of neurological disorders. *Nat. Rev. Neurosci.* 15, 683–695.
5. Sporns, O. (2014). Contributions and challenges for network models in cognitive neuroscience. *Nat. Neurosci.* 17, 652–660.
6. Gazzaniga, M. S., ed. (2009). *The Cognitive Neurosciences*. MIT Press.

7. Szatmary, B. and Izhikevich, E. M. (2010). Spike-timing theory of working memory. *PLoS Comp. Bio.* *6*, e1000879.
8. Feldt, S. *et al.* (2011). Dissecting functional connectivity of cortical microcircuits: experimental and theoretical insights. *Trends Neurosci* *34*, 225–236.
9. Curto, C. and Itskov, V. (2008). Cell groups reveal structure of stimulus space. *PLoS Comp. Bio.* *4*, e1000205.
10. Crossley, N. A. *et al.* (2013). Cognitive relevance of the community structure of the human brain functional coactivation network. *Proc. Natl. Acad. Sci. USA* *110*, 11583–11588.
11. Bullmore, E. and Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* *10*, 186–198.
12. Ghrist, R. (2014). *Elementary Applied Topology*. 1.0 edition, Createspace.
13. Nanda, V. and Sazdanović, R. (2014). Simplicial models and topological inference in biological systems. In *Discrete and Topological Models in Molecular Biology* pp. 109–141. Springer.
14. Kozlov, D. (2007). *Combinatorial Algebraic Topology*, vol. 21,. Springer Science & Business Media.
15. Giusti, C. *et al.* (2015). Clique topology reveals intrinsic geometric structure in neural correlations. *Proc. Nat. Acad. Sci. USA* *112*, 13455–13460.
16. Kim, E. *et al.* (2014). Morphological brain network assessed using graph theory and network filtration in deaf adults. *Hearing Res.* *315*, 88–98.
17. Arai, M. *et al.* (2014). The effects of theta precession on spatial learning and simplicial complex dynamics in a topological model of the hippocampal spatial map. *PLoS Comp. Bio.* *10*(6), e1003651.
18. Dabaghian, Y. *et al.* (2012). A topological paradigm for hippocampal spatial map formation using persistent homology. *PLoS Comp. Bio.* *8*, e1002581.
19. Dabaghian, Y. *et al.* (2014). Reconceiving the hippocampal map as a topological template. *Elife* *3*, e03476.
20. Chen, Z. *et al.* (2014). Neural representation of spatial topology in the rodent hippocampus. *Neural Comput.* *26*, 1–39.
21. Ellis, S. and Klein, A. (2014). Describing high-order statistical dependence using “concurrence topology” with application to functional MRI brain data. *H.H.A.* *16*(1), 245–264.

22. Bassett, D. S. *et al.* (2015). Learning-induced autonomy of sensorimotor systems. *Nat. Neurosci.* *18*, 744–751.
23. Pirino, V. *et al.* (2014). A topological study of repetitive co-activation networks in in vitro cortical assemblies. *Phys. Bio.* *12*, 016007–016007.
24. Singh, G. *et al.* (2008). Topological analysis of population activity in visual cortex. *J. Vis.* *8*, 11.11118.
25. Chung, M. K. *et al.* (2009). Persistence diagrams of cortical surface data. In *Information Processing in Medical Imaging* pp. 386–397, Springer.
26. Brown, J. and Gedeon, T. (2012). Structure of the afferent terminals in terminal ganglion of a cricket and persistent homology. *PLoS ONE* *7*, e37278.
27. Bendich, P. *et al.* (2014). Persistent homology analysis of brain artery trees. *Ann. Appl. Stat.* *to appear*.
28. Lee, H. *et al.* (2011). Discriminative persistent homology of brain networks. In *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on* pp. 841–844, IEEE.
29. Khalid, A. *et al.* (2014). Tracing the evolution of multi-scale functional networks in a mouse model of depression using persistent brain network homology. *NeuroImage* *101*, 351–363.
30. Choi, H. *et al.* *et al.* (2014). Abnormal metabolic connectivity in the pilocarpine-induced epilepsy rat model: a multiscale network analysis based on persistent homology. *NeuroImage* *99*, 226–236.
31. Petri, G. *et al.* (2014). Homological scaffolds of brain functional networks. *J. Roy. Soc. Int.* *11*, 20140873.
32. Nicolau, M. *et al.* (2011). Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. *Proc. Nat. Acad. Sci. USA* *108*, 7265–7270.
33. Gameiro, M. *et al.* (2013). A topological measurement of protein compressibility. *Jpn. J. Ind. Appl. Math.* *32*, 1–17.
34. Xia, K. *et al.* (2015). Persistent homology for the quantitative prediction of fullerene stability. *J. Comp. Chem.* *36*, 408–422.
35. Chan, J. M. *et al.* (2013). Topology of viral evolution. *Proc. Nat. Acad. Sci. USA* *110*, 18566–18571.
36. Boczko, E. M. *et al.* (2005). Structure theorems and the dynamics of nitrogen catabolite repression in yeast. *Proc. Nat. Acad. Sci. USA* *102*, 5647–5652
37. Carlsson, G. (2009). Topology and data. *Bull. Amer. Math. Soc.* *46*, 255–308.

38. Lum, P. *et al.* (2013). Extracting insights from the shape of complex data using topology. *Sci. Rep.* *3*.
39. Dowker, C. H. (1952). Homology groups of relations. *Ann. Math.* *56(1)*, 84–95.
40. Fortunato, S. (2010). Community detection in graphs. *Phys. Rep.* *486*, 75–174.
41. Porter, M. A. *et al.* (2009). Communities in Networks. *Not. Amer. Math. Soc.* *56*, 1082–1097, 1164–1166.
42. Bassett, D. S. *et al.* (2014). Cross-linked structure of network evolution. *Chaos* *24*, 013112.
43. Bassett, D. S. *et al.* (2011). Dynamic reconfiguration of human brain networks during learning. *Proc. Natl. Acad. Sci. USA* *108*, 7641–7646.
44. Bassett, D. S. *et al.* (2013). Task-Based Core-Periphery Structure of Human Brain Dynamics. *PLoS Comp. Bio.* *9*, e1003171.
45. Achard, S. *et al.* (2006). A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *J. Neurosci.* *26*, 63–72.
46. Ginestet, C. E. *et al.* (2011). Brain network analysis: separating cost from topology using cost-integration. *PLoS One* *6*, e21570.
47. Bassett, D. S. *et al.* (2012). Altered resting state complexity in schizophrenia. *NeuroImage* *59*, 2196–207.
48. Garrison, K. A. *et al.* (2015). The (in)stability of functional brain network measures across thresholds. *NeuroImage* *S1053-8119*, 00428–0.
49. Drakesmith, M. *et al.* (2015). Overcoming the effects of false positives and threshold bias in graph theoretical analyses of neuroimaging data. *NeuroImage* *118*, 313–333.
50. Sala, S. *et al.* (2014). pFDR and pFNR estimation for brain networks construction. *Stat. Med.* *33*, 158–169.
51. Langer, N. *et al.* (2013). The problem of thresholding in small-world network analysis. *PLoS One* *8*, e53199.
52. Rubinov, M. and Bassett, D. S. (2011). Emerging evidence of connectomic abnormalities in schizophrenia. *J. Neurosci.* *31*, 6263–5.
53. Grosenick, L. *et al.* (2015). Closed-loop and activity-guided optogenetic control. *Neuron* *86*, 106–139.
54. Krug, K. *et al.* (2015). Understanding the brain by controlling neural activity. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* *370*, 20140201.

55. Sulzer, J. *et al.* (2013). Real-time fMRI neurofeedback: progress and challenges. *NeuroImage* 76, 386–399.
56. Taylor, D. *et al.* (2015). Topological data analysis of contagion maps for examining spreading processes on networks. *Nat. Comm.* 6. 7723
57. Raj, A. *et al.* (2012). A network diffusion model of disease progression in dementia. *Neuron* 73, 1204–1215.
58. Zhou, J. *et al.* (2012). Predicting regional neurodegeneration from the healthy brain functional connectome. *Neuron* 73, 1216–1227.
59. Carlsson, G. *et al.* (2008). On the local behavior of spaces of natural images. *Int. J. Comput. Vis.* 76, 1–12.