FRONT PAGE

PROJECT FINAL REPORT

Grant Agreement number:

Project acronym:

Project title:

Funding Scheme:

Period covered: from to

Name of the scientific representative of the project's co-ordinator[1], Title and Organisation:

Tel:

Fax:

E-mail:

Project website Error! Bookmark not defined. **address:**

[1] Usually the contact person of the coordinator as specified in Art. 8.1. of the Grant Agreement.

4.1 Final publishable summary report

Executive Summary

The goal of the TOPOSYS project is to build a theoretically sound foundation for the study of dynamic multi-scale complex systems using newly developed tools in applied and computational topology. Computational topology is a relatively new and rapidly developing field, which extends tools from algebraic topology to more computationally friendly forms. It also deals with extending the corresponding theory to situations, which are more applicable to scientific inquiry (such as defining a meaningful topology of a sampled metric spaces).

At their core, these tools are well-suited to the study of complex systems – they have a intrinsic local-to-global flavour, extracting and inferring qualitative global behaviour from local observations. The ability to go from local to global scales is crucial to develop the rigorous notions of multi-scale behaviour. Likewise, the qualitative, yet mathematically well-founded nature of the descriptions of systems abstract away the details which allows for a clearer analysis. For example, in the study of dynamical systems –we cannot describe the path a particular flow will take around a strange attractor in a chaotic system, we can often say meaningful things about the trajectory as an entirety, and its abstract properties.

The project has advanced the state-of-the-art on many different fronts including the persistence of maps, the connection between persistence and statistics, various generalizations of persistence (and its application to dynamical systems) as well as applications. There have been a number of papers in top journals and conferences. There have also been surprising connections, which have led to initial work in semigroups, knot theory, and other invariants such as the Euler integral and curve and the fundamental group. The only area where there was the least progress was in the examination of social media data. The lack of progress was the result of a lack of a topological signal in the data. This however led us to the study of topology (and persistence) on Bregman divergences which ties together information geometry and applied topology, opening the way for a whole new area of investigation beyond the project.

In this report we give an overview of the main results of the project. The descriptions are intentionally terse - with pointers to the corresponding papers. We have tried to group the results according to where they fit within the project - but in many instances, some results, particularly theorems, will have undefined quantities. We will try to explain these informally, however, the corresponding papers should be considered the final authority.

Overview of the project

The focus of the project is on techniques based on persistence, an inherently multi-scale approach. Rather than study a single scale or time slice, persistence studies is how a space changes either over different scales or over time. Furthermore, the approach is algorithmic with a strong emphasis on computable invariants – especially for large scale and high dimensional data. With the algorithmic approach, we are able to consider inverse problems, such as reconstructing dynamical behaviours from discrete point samples.

Overview of Persistence

To place things into perspective, we include an overview of persistence which is a recurring tool we use.

The study of persistent homology originates from Edelsbrunner, Letscher, and Zomorodian, who first define the term and provide an algorithm for the computation of persistent homology. Taking their inspiration from α-shapes, the authors assume that a filtered simplicial complex is provided as input, and produce a description of its persistent homology. In a slightly later paper, Edelsbrunner, Harer, and Zomorodian demonstrate that persistent homology can be applied to morse complexes from piecewise linear functions on a manifold – the filtered simplicial complex required is given by combining the morse complex cells with the function values at the critical points witnessing each cell.

From this point and onwards, one strongly present culture in persistent homology remains focused on the role of a function defined on a manifold as the input data for the method. This viewpoint has proven remarkably fruitful in the study of stability, and provides the best tools we currently have for justifying topological inferences with persistent homology. The functional perspective is shown in Figure 1.

It is worth noting that a point cloud topology point of view fits in this framework: as is illustrated in Figure 2, the distance to a discrete set of points produces a real-valued function on the ambient space of the points, with a persistent homology corresponding closely to the Cech complex homology of the point cloud itself.

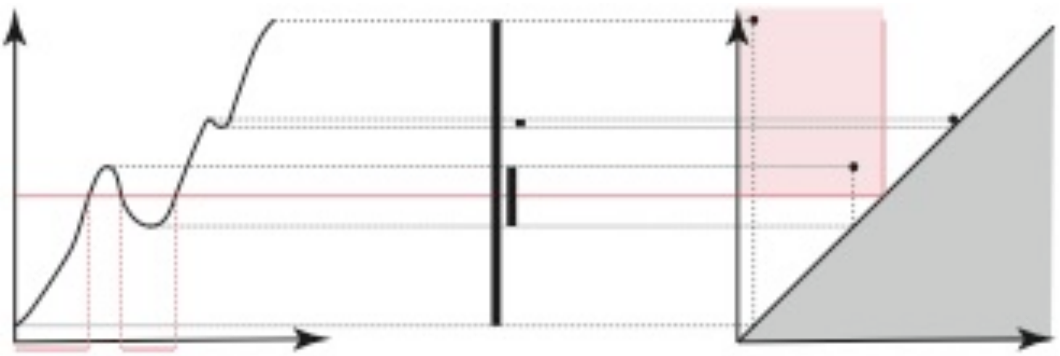

Figure 1. 0-dim Persistence of sublevel sets of a function. In black, we see the three components that appear at different times show up – in the middle in a persistence barcode, and to the right as the three points in a persistence diagram. In red, we indicate a particular choice of height ε, at which the sublevel set has two components – drawn below the graph to the left. These two components can be read off in both persistence visualizations – through the two intersected bars in the middle, and through the two points contained in the shaded red region to the right.

Here we also see the two different versions of representing persistence: barcodes and persistence diagrams. We use them interchangably as they are equivalent. This is the tip of the iceberg as the categorical and algebraic viewpoints have become more prominent during the time of the project. See ``Sketches of a platypus: A Survey of Persistent Homology and its Algebraic Foundations" (http://arxiv.org/abs/1212.5398) for a complete overview of the different viewpoints to persistence. It is worth noting that this project has introduced a new topos theoretic perspective.

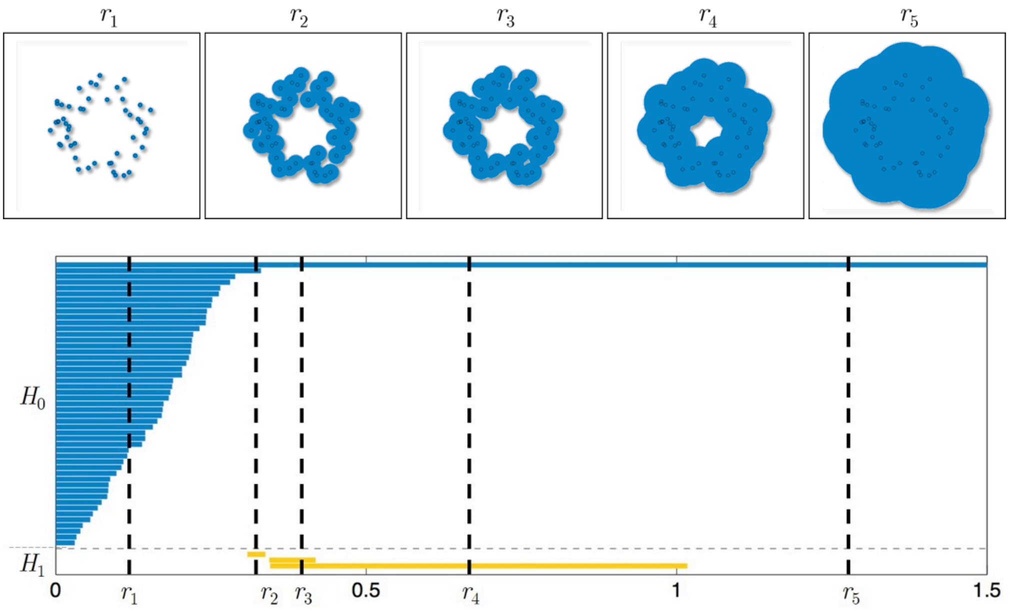

Figure 2 (top) a union of balls of radius r around a random set of n =50 points, uniformly distributed on an annulus in R2. We present five snapshots of this filtration. (bottom) The persistent homology of the filtration. The x-axis is the radius r, and the bars represent the cycles that born and die.

In addition to introducing generalizations and variants of persistence (and the corresponding algorithms to compute them), a recurring theme is **stability**. The stability of persistence diagrams is one of the key results in applied topology and we highlight where our results have corresponding stability guarantees.

Objectives: The objectives of this project align along three main directions:

1. **Persistence of maps** – The first goal is to move to a more explicitly *functorial* approach. Rather than concentrate on the objects (e.g. sampled spaces), we concentrate on the maps between these objects. For example, a dynamic system can be viewed as a map from a space to itself, where each application of the map iterates the system forward. To understand the dynamics of a system, we must understand the properties of this map. The first objective within this goal is to define a theory of persistence on the space of maps, which is amenable to sampled data. The second objective is to extract more refined information in cases where this is possible. For example, a natural model recurrent dynamics is the circle, which can be parameterized by cohomological techniques. Finally, in complex systems, we often have vastly different measurements, which must be combined coherently which we examine as the problem of exploring different types of hierarchies of maps.

2. **Statistics and topology** – Statistics forms the basis of data analysis, but it has proven difficult to apply it in topological settings. One of our goals is to help bridge the gap between statistics and topology – using the strengths of both. The first objective is to gain a better

understanding of the statistics of topological noise. Topological results give a bound on the magnitude of the noise but relies on quantites which cannot be measured or estimated. More advanced probabilistic techniques will allow more general noise models to be considered and gaining a more complete understanding the behavior of the noise in our techniques. Furthermore, the exploration of the consequences of doing statistics on the output of persistence theory is still relatively unexplored. The existing results point to the possibility of rigorous model selection – which is our second objective. Our final objective is to combine topology and statistics to better understand smoothing processes as this is a natural component in many multi-scale notions.

3. **Categorical foundations -** Category theory is a powerful language, which is already ubiquitous in mathematics. Phrasing existing techniques within a categorical setting often reveals useful abstractions and connections. Our goal is to explore categories as a fundamental abstraction tool for topological and algebraic approaches to systems. The first objective is to place persistence in the setting of a topos of sheaves, which if successful will unify various aspects of the foundations of persistence theory. Our second objective is to examine the use of sheaves in a statistical or machine learning context. Sheaf theory is a formal machinery which allows for general transitions between local and global properties. This has gained intense interest in the applied topology community over the last few years. Finally, we look at generalizations of persistence so that they may be more directly relevant to dynamic systems.

4. **Validation and Applications** – to validate the efficacy of our approach, the objective is to validate with applications from robotics and social media. The former is not a traditional "complex systems," however serves as an important stepping stone towards more complex systems. Social media refers primarily to textual linked data, which may come from Twitter, Wikipedia, or other sources. Note that the methods will not be tied to these datasets and as opportunities arise, other datasets will be considered.

The project combines four main areas of mathematics: category theory, statistics, and dynamical systems with computational topology as the joint platform for the three other component in order to work towards a mathematically rigorous description of the dynamics of a system from a local to a global scale. In this framework, multi-scale features have a natural place, and the focus on computation and algorithmics, reflecting on both the theoretical and practical sides of computer science, led to verification of the developed theory. In our work, we also ended up using topos theory, lattice theory, integral geometry, discrete Morse theory, just to name a few.

Main S&T results

Persistence of Self-Maps

One of the successes of the project is the study of the persistence of endomorphisms induced in homology by continuous self-maps. In the case of finitely generated homology with field coefficients, the induced homomorphism by a continuous map between topological spaces is a linear map between finite-dimensional vector spaces. Such a map

$$\varphi : Y \to X$$

is characterized up to conjugacy by its rank. This is in contrast to a linear self-map,

$$\phi : X \to X$$

which in the case of an algebraically closed field is characterized up to conjugacy by its Jordan form.

Initially, we have developed an algorithm for computing the persistent eigenspaces for a fixed eigenvalue. Beyond the required theoretical framework, called towers of vector spaces (which are a categorical notion discussed later) and the algorithm, we have proved that the persistence diagram it produces is stable under perturbations of the input, and the algorithm converges to the homology of the studied map, reaching the correct ranks for a sufficiently fine sample. We have shown an **inference theorem** is in line with stability results for persistence diagrams, which are given in terms of Hausdorff distance. The theorem states roughly, that given some conditions if we approximate the graph of a map, denoted above by G, then we obtain the correct rank of the homological eigenspaces. The idea is illustrated by a toy example, shown in Figure 3. Consider the circle in the complex plane, and let the map double the angle of an input point. The doubling can only be seen at a certain scale in the second to last pane. For the map to be well defined, the image of a simplex must have already entered the filtration, which again limits the resolution at which we can examine the map.

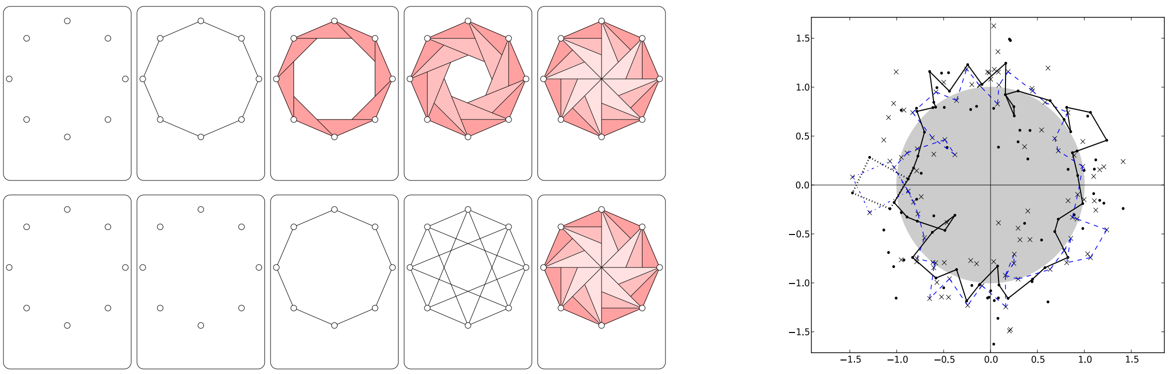

Figure 3 Five simplicial complexes in the filtration of the eight data points at the top, and the domains of the induced partial simplicial maps at the bottom. On the right we see the sampled map.

One of the drawbacks of this work is scalability. The original algorithm depends on repeated computation of the Smith Normal Form. We have since extended the above algorithm into an incremental algorithm based on algebraic insights, which promises to speed up the computation, with the paper in preparation.

Another more general problem with the initial algorithm was to compute the persistence of a self-map, we must first determine the eigenvalues. That is, for two maps, we must find all t's such that

$$\varphi - t\psi = 0$$

has a solution.

It may happen over a finite field that every element of the field is an eigenvalue, a phenomenon which is called "an abundance of eigenvalues." This difficulty in finding the eigenvalues for the pair of maps and in identifying them as dynamically significant, raised the question whether there exists a way to build the sequence of eigenspaces, and compute their persistence, for all eigenvalues simultaneously. The correct mathematical structure to study this question is the Kronecker canonical form for *m x n* matrix pencils, $tB - A$. We have developed the corresponding techniques to compute this.

We have also developed an incremental algorithm to improve scalability along with the techniques described in the next section.

Cech and Delaunay Complexes

A fundamental task in the study of the self-map as well as topological data analysis is to turn the data into a topological space, or a filtration of spaces. Assuming the data consists of points in Euclidean space, we have several possible ways to construct a filtration using a non-negative scale parameter *r*. We can connect *p+1* data points by the spanned *p*-simplex if and only if the closed balls of radius *r* centered at the points have a non-empty common intersection. This gives the Cech complex for radius *r*, which is known to have the same homotopy type as the union of balls (and hence the same homology type). It was also previously known that certain subcomplexes of the Delaunay-Cech complex, where the balls are intersected by the Voronoi regions around the points, also shares this property. We have shown that these complexes are simple-homotopy equivalent – that is there exists a sequence of simple homotopies from one complex to the other. This also extends to another complex called the Wrap complex as well as to weighted versions.

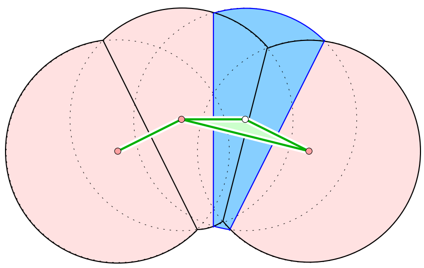

Figure 4 The Voronoi balls of four points in the plane, the shaded three of which impose constraints. The corresponding selective Delaunay complex has four edges and one triangle. It properly contains the Delaunay complex and is properly contained in the Cech complex of the four points.

These Delaunay-Cech has the property that it is much smaller than just the Cech complex as well as having other nice geometric properties. This enables the computation to scale to a larger number of input points, since there are fewer higher dimensional simplices, as well as extending this to the representation of chain maps.

Cohomology of Recurrent Systems

Recurrence, as well as periodicity, play a central role in characterizing the behavior of dynamic systems. The first looks at characterizing recurrence by building a topological model of it – a closed one form embedded in a high dimensional space. The second looks at the problems associated with computing the Conley index of Poincare maps.

Circle Valued Analysis of Recurrence

There are two key observations which lead to this approach. The first is a technique of Takens:

> **Definition:** Given a time-series $f : t \to \mathbb{R}$, a **time-delay embedding** is a lift to a time-series $\phi : t \to \mathbb{R}^d$ defined by

$$\phi(t) = (f(t), f(t+\alpha), \ldots, f(t+(d-1)\alpha))$$

Takens' proved a remarkable theorem about these embeddings. The theorem states that almost every time-delay embedding of a 1-dimensional measurement (time-series) can recover the underlying manifold and the dynamics of the system. We have developed a pipeline shown in Figure 5 to use this technique to analyse dynamical systems

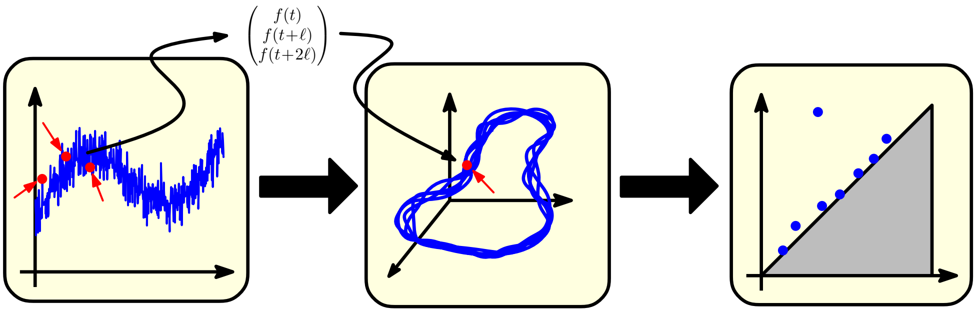

Figure 5 The first step is to choose parameters and compute a time-delay embedding. The persistence diagram of a distance filtration gives a natural measure of the quality of the embedding based on which we can change the time-delay parameters.

For example, the simplest periodic system is a sinusoid, which precisely traces out a circle in phase space. The circle is not present in the 1-dimensional signal, but appears upon boosting to 2. We describe three types of behavior along with their corresponding topological models:

- Periodic systems correspond to a circle.
- Quasiperiodic systems are modeled by a k-dimensional torus. For instance a system with two periods will trace out a trajectory on a torus T^2. Depending on the ratio of the periods, the trajectory will be a dense curve on the torus or a torus knot. Circular coordinates on the torus can be exploited to extract this periodicity information, in a sort of 'topological Fourier analysis.'
- Recurrent systems are modeled by a bouquet of circles (also known as a rose). The archetypal example is the Lorenz system. Its dynamics circulate repeatedly around the two wings of a butterfly-shaped attractor in an aperiodic sequence.

In computing these topological models, persistence can also help us choose parameters, by trying different parameters and finding which have the "cleanest" topological signal (i.e. the most persistent

1-dimensional cohomology class). Once the parameterization has been computed, we can perform a number of analyses. We present two here, finding the period and tracking which attractor we are circling in the Lorenz attractor.

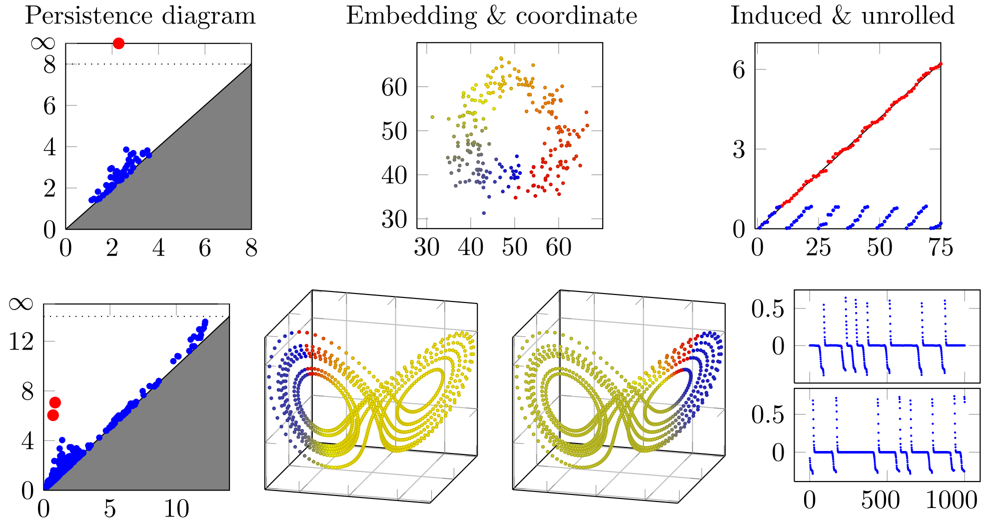

Figure 6 (Top) Monthly rainfall measurements from Nottingham. On the right, we unroll the coordinate and fit a linear model. Computing the slope in the unrolled coordinate yields 1{12.002 with a correlation of 0.9995. (Bottom) The Lorenz attractor. To the left, the persistence diagram from the point cloud we computed, with the chosen cocycles marked in large red dots. The resulting coordinates are indicated left to right, with their corresponding coordinate functions on the time series top to bottom in the right-most plot. The recurrent structure can be clearly seen in the pulse behaviour at the far right

Later we also describe an application of these methods to robotics for gait analysis.

Poincare Maps

Standard techniques in computational dynamics often require difficult and low rigorous integration. A natural step is to apply an analogous approach for the computation of the Conley index of a Poincare map. A Poincare map, also known as a first recurrence map, looks at how the image of a recurrent map looks when it has "returned." Computing the Conley index provides important information on the behaviour of a dynamical system. Algorithms based on this then do not require a long time integration along the trajectories of the flow and benefit from the general algorithms computing index pairs for discrete dynamical systems.

We developed the algorithms and theory behind this approach. The theoretical side turned out to be surprisingly complicated, however, the algorithms based on this approach work as expected.

We have also worked on how best to perform the discretization to do this computation, yielding in rigorous but more scalable results.

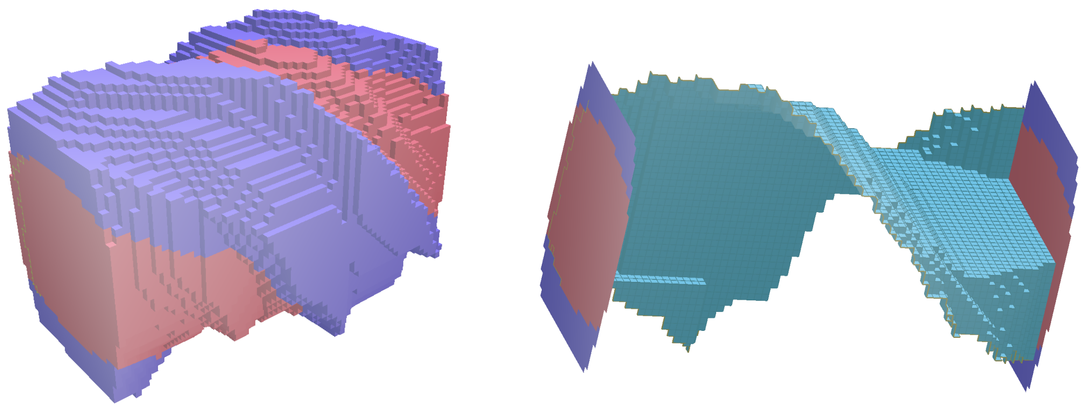

Figure 7 The left figure shows the full cubes in a pair of spaces and a representing 1-chain of a generator. The right figure shows the constructed 2-chain in cyan. The constructed 1-chain can be seen at the right edge. It is the image under the endomorphism in homology induced by the index map of the Poincare map.

Multiscale Statistics

Crackle

In data there is often oncertainty, so we also looked at the homology of simplicial complexes built on a random set of vertices. Depending on the randomness that generates the vertices, the homology of these complexes can either become trivial as the sample size grows, or can contain more and more complex structures. The motivation for these results comes from applications of topological tools for pattern analysis, object identification, and especially for the analysis of data sets. Typically, one starts with a collection of points and forms some simplicial complexes associated to these, and then takes their homology.

An important example occurs in the following manifold learning problem: Let M be an unknown manifold embedded in a Euclidean space, and suppose that we are a given a set of independent and identically distributed (iid) random samples from the manifold. In order to recover the homology of M, we consider the homology of the union of balls of radius *r* around the points. If the number of sample points is "large enough," the homology of the union of balls will be the same as the underlying manifold (for appropriately chosen r).A confounding issue arises when the sample points do not necessarily lie on the manifold, but rather are perturbed from it by a random amount. When this happens, it will follow from our results that the precise distribution behind the randomness plays a qualitatively important role. It is known that if the perturbations come from a bounded or strongly concentrated distribution, then they do not lead to much spurious homology, and the above line of attack, appropriately applied, works.

Figure 8 provides an illustrative example of what happens when sampling points from an annulus and perturbing them with additional noise before reconstructing the annulus. In particular, it shows that if the additional noise is in some sense large then sample points can appear basically anywhere, introducing extraneous homology elements.

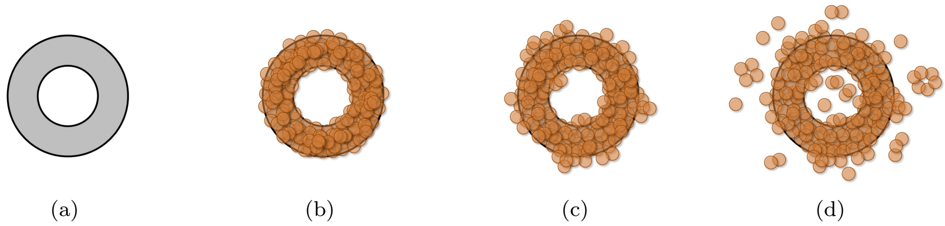

Figure 8 (a) The original space M (an annulus) that we wish to recover from random samples. (b) With the appropriate choice of radius, we can easily recover the homology of the original space from random samples from M. (c) In the presence of bounded noise, homology recovery is undamaged. (d) In the presence of unbounded noise, many extraneous homology elements appear, and significantly interfere with homology recovery.

To study these phenomena, we considered three representative examples of probability densities. These are the power-law, exponential, and the standard Gaussian distributions. For large samples from any of these distributions we shall show that there exists a 'core' - a region in which the density of points is very high and so placing unit balls around them completely covers the region. Consequently, the Cech complex inside the core is contractible. The size of the core obviously grows to infinity as the sample size n goes to infinity, but its exact size will depend on the underlying distribution. For the three examples above, if we denote the radius of the core by R_n^c , we proved that

$$R_n^c \sim \begin{cases} (n/\log n)^{1/\alpha} & f(x) \propto \frac{1}{1+|x|^\alpha}, \\ \log n & f(x) \propto e^{-|x|}, \\ \sqrt{2\log n} & f(x) \propto e^{-|x|^2/2}. \end{cases}$$

Note that in all three cases we have tacitly assumed that the cores are balls, a natural consequence of the spherical symmetry of the probability densities. Beyond the core, the topology is more varied. For fixed n, there may be additional isolated components, but no longer enough placed densely enough to connect with one another and to form a contractible set. Indeed, we shall show that the individual components will typically have enough homology to be, individually, non-contractible. Thus, in this region, the topology of the Cech complex is highly nontrivial, and many homology elements of different orders appear. We call this phenomenon 'crackling' and is shown in the Figure below

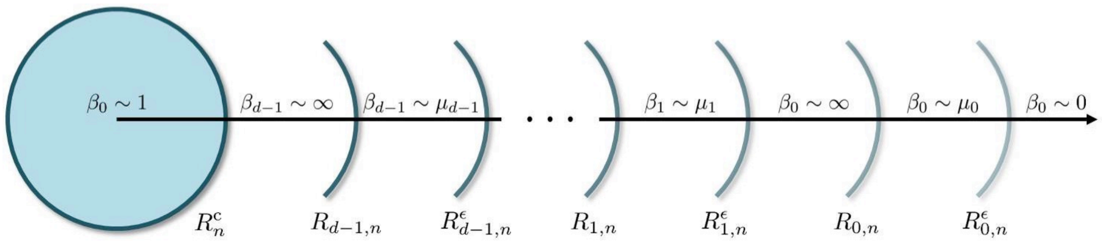

Figure 9 The layered behaviour of crackle. Inside the core the complex consists of a single component and no cycles. The exterior of the core is divided into separate annuli. Going from right to left, we see how the Betti numbers grow. In each annulus we present the Betti number that was most recently changed.

We have also experimentally validated these results which was presented at ECCS 2014.

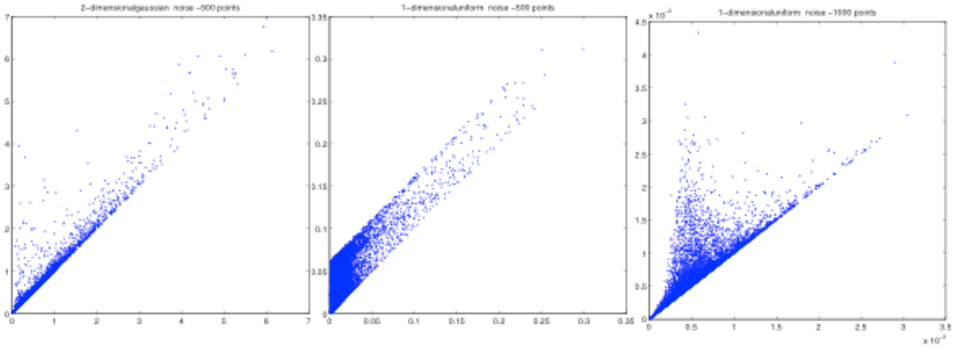

Figure 10 Persistence diagrams of different models of noise

Topological invariants of random processes

There has also been extensive work on understanding Betti numbers and the Euler integral of random processes under different conditions. In particular, there are central limit theorems for the Betti numbers for **general stationary processes** as well as for stationary Poisson processes and concentration results for inhomogeneous Poisson and binomial point processes. We also have results (including central limit theorems) in the so-called "thermodynamic" regime (which includes the percolation threshold) in which the complexes become very large and complicated, with complex homology characterised by diverging Betti numbers. These all help us better understand the topology of random spaces.

In addition to stationary point processes, we have results on the dynamic Erdos-Renyi graph, where edges enter and leave the complex over time. We have results showing that Betti numbers converge to a stationary Ornstein-Uhlenbeck process.

Our interest was not only limited to Betti numbers as we also have central limit theorems for the Euler integral of a Gaussian random field. Finally, we have extended the results on Betti numbers in the thermodynamic regime to **persistent Betti numbers**, proving a central limit theorem.

Maximal persistence

We initiate the study of persistent homology of random geometric simplicial complexes. Our main interest is in maximally persistent cycles of degree-k in persistent homology, for a either the Cech or the Vietoris-Rips filtration built on a uniform Poisson process of intensity n in the unit cube. This is a natural way of measuring the largest ``k-dimensional hole'' in a random point set. We show that for all $d \geq 2$ and $1 \leq k \leq d-1$ the maximally persistent cycle has persistence of order

$$\Theta\left(\left(\frac{\log n}{\log\log n}\right)^{1/k}\right),$$

with high probability, characterizing its rate of growth as $n \to \infty$. Below, in Figure 11, we have an example of the advantage of multiplicative persistence. The set P consists of all the points in the figure, and we look at the distance filtration. The persistent homology includes two cycles one generated by the corners of an equilateral triangle of side 2, and one generated by 16 points arranged on a circle of radius 0.1. Intuitively, the second seems to be a real feature of the data, while the first seems to be an artefact generated by outliers. However, using death-birth the first is more persistent while using death/birth the latter is more persistent.

This was also validated in experiments, which suggest that a central limit theorem could be proven about multiplicative persistence.

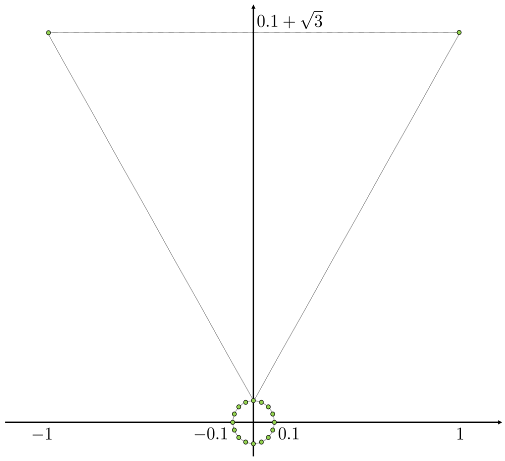

Figure 11 An example for multiplicative persistence

Multi-scale Kernels

Visual data is typically piped through complex processing chains in order to extract information that can be used to solve high-level inference problems, such as recognition, detection or segmentation. This information might be in the form of low-level appearance descriptors, e.g., SIFT or of higher-level nature, e.g., responses of batteries of object detectors or features extracted at specific layers of deep convolutional networks. In many problems, the consolidated data is then fed to some discriminant classifier such as the popular SVMs. In this we work we looked at how to use persistence diagrams as features in machine learning algorithms

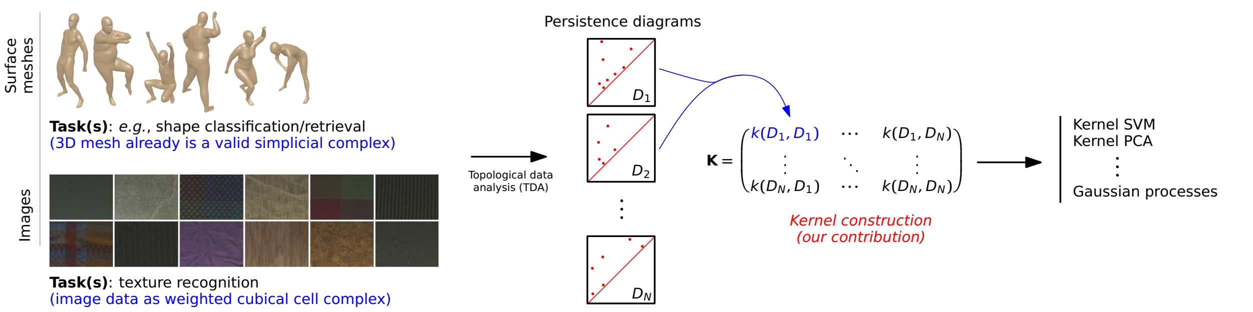

Figure 12 The pipeline for analysis using persistence diagrams including where the kernel construction fits in.

We developed a multi-scale L2 embedding of persistence diagrams, based on the principles of heat diffusion. The L2 norm between embedded PDs defines a distance measure that is bounded, from above, by the degree-1 Wasserstein distance. We showed that the negative of this squared norm is a valid (conditionally) positive definite kernel and can thus be used within most kernel-based learning techniques. The heat parameter of our kernel controls its robustness to noise and can be tuned to the data (just like the parameters of a RBF kernel).

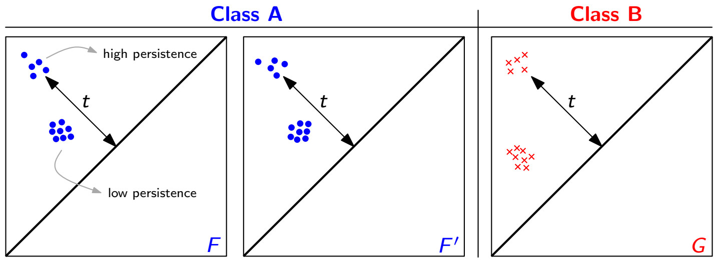

Figure 13 Two instances of persistence diagrams F,F' class A and one diagram G from class B}. The classes only differ in their points of low-persistence (ie, points closer to the diagonal).

From a conceptual point of view, Bubenik's concept of persistence landscapes is possibly the closest to ours, since an embedding of persistence diagrams is proposed. While persistence landscapes were not explicitly designed for use in machine learning algorithms, we show that they in fact admit the definition of a valid (conditionally) positive definite kernel. In summary, persistence landscapes as well as our approach represent computationally attractive alternatives to the bottleneck or Wasserstein distance.

This technique was tested on shape retrieval and shape classification with positive results.

Simplification of 2 and 3 dimensional vector fields

Vector field simplification aims to reduce the complexity of the flow by removing features in order of their relevance and importance, to reveal prominent behavior and obtain a compact representation for interpretation. Most existing simplification techniques based on the topological skeleton successively remove pairs of critical points connected by separatrices, using distance or area-based relevance measures. These methods rely on the stable extraction of the topological skeleton, which can be difficult due to instability in numerical integration, especially when processing highly rotational flows.

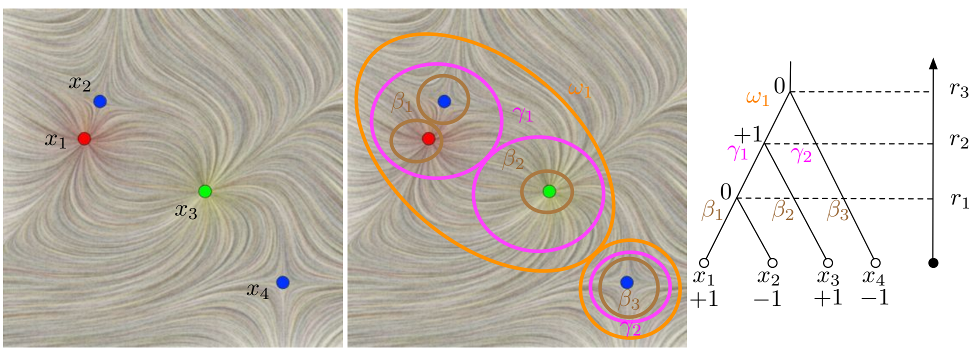

Figure 14 Suppose the vector field is continuous, where sinks are red, sources are green, and saddles are blue. From left to right: vector fields, the relations among components of the sublevel sets, and augmented merge trees.

We have developed a new vector field simplification scheme derived from the recently introduced notion of robustness. Robustness, a notion related to persistence is used to represent the stability of critical points and assess their significance with respect to perturbations of the vector field. Intuitively, the robustness of a critical point is the minimum amount of perturbation, with respect to a metric encoding flow magnitude that is required to cancel it within a local neighbourhood. Based on this we developed a new hierarchical simplification strategy based on robustness, which enables the pruning of sets of critical points according to a quantitative measure of their stability. In a series of paper, we also extended this algorithm to the time-varying case as well as to the three dimensional case. Interestingly, we found that even in three dimensions, we found we could remove critical points from vector fields with relatively small perturbations but this did not simplify the vector field visually as the flow remained complicated.

Laplacian Clustering

We have developed a clustering algorithm, which combines the topological information with spectral clustering. For input sampled from a uniform distribution supported on a metric space *X*, it outputs a clustering of the data based on a topological estimate of the connected components of *X*. The algorithm works by choosing a weighted graph on the samples from a natural one-parameter family of graphs using an error based on the heat operator on the graphs. The estimated connected components of *X* are identified as the support of the eigenfunctions of the heat operator with eigenvalue 1, which allows the algorithm to work without requiring the number of expected clusters as input.

The algorithm directly uses the topological information in the Laplacian and heat operator, and which also demonstrates the utility of considering clustering as a problem of estimating the number of connected components of a distribution whose support is disconnected. Second, the algorithm produces both the number of clusters and the clusters themselves. The algorithm we present also gives, we believe, the first use of cross-validation in a non-commutative context. While this appears formally like standard cross-validation, we find ourselves outside the standard context of commutative probability, and the usual proofs of convergence no longer apply. Rigorous justification of this cross-validation technique is an important topic for future study, but we do not address this here. Finally, we give the first algorithm for automatically choosing the bandwidth of the kernel function used in approximations to Laplace-Beltrami operators.

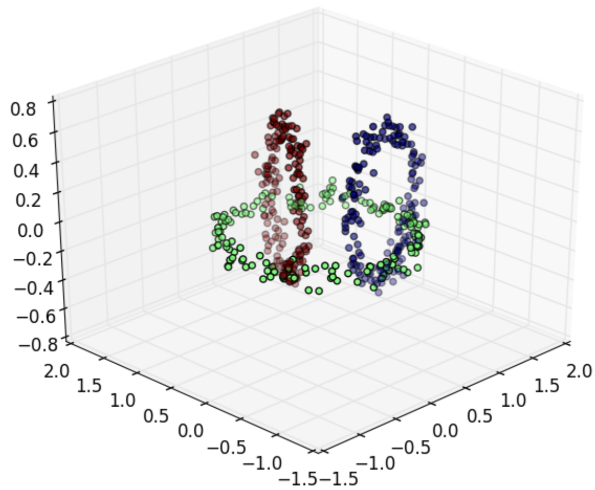

Figure 15 Clusters found using the Laplacian technique

Categorical Systems

Topoidal Foundation of Persistence

Another long standing goal of the project was a generalized persistence theory using the topos theory as a foundation for persistence. The main idea is to replace the notion sublevel sets of the real line or natural numbers (as in one dimensional or classical persistence) or higher dimensional analogues of sublevel sets (which consist of cones or unions of cones) with the general notion of a Heyting algebra. This can be thought of as a bounded distributed lattice – the prototypical example being open sets on a topological space. One of the basic observations of topos theory is that we can replace the notion of sheaves over open sets with sheaves over a Heyting algebra. The lattice operations of meet and join mimic the operations of intersection and union allowing us to formulate the (co)-sheaf condition. This generality allows us to encompass existing persistence theories (e.g,. classical, multidimensional, zig-zag, etc). In the initial stages of the project, the concentration was on identification of suitable Heyting algebras, with our initial test case being zig-zag persistence. The resulting algebra of bars is shown below in Figure 16.

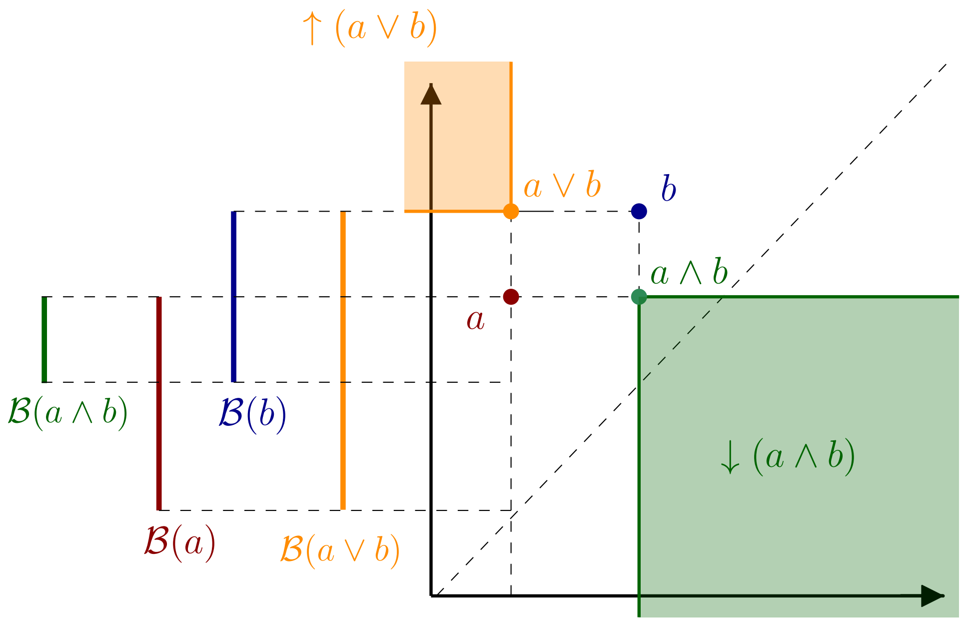

Figure 16 The algebra of bars for zig-zag persistence. This is the algebra of intervals – interestingly we must allow for negative intervals to enure this is a Heyting algebra.

Once we have defined this Heyting algebra, which we refer to as an algebra of lifetime intervals P, we defined a variable-time set theory (which is a topos over this algebra). This can be thought of as a data structure, where for any possible lifetime we return its corresponding set. In the one dimensional case this take the form of all points in the space which have a filtration value less than x. Based on this new set theory, we developed a generic algorithm for computing homology – which gives us a persistence module (which is sheaf valued). The algorithm is based on computing homology pointwise (over each possible lifetime) as a presheaf and then sheafifying, which can be done by taking colimits (i.e. computing the stalks) and then taking limits (i.e. finding global sections).

This construction is completely general and allows us to generate new persistence theories depending on the “shape” we choose. It appears as the extra structure of a Heyting algebra compared to a general partially ordered set (POSET) prevents several examples of what can go wrong with other sheaf-based approaches. Finally, we have established many of the ingredients required to prove a bottleneck matching type results for the indecomposables of such persistence modules. There are many directions, which remain open. One is the algorithnmic aspect – the nature of a Heyting algebra gives us exactness which suggests a natural approximation scheme which is important since an algebra of lifetimes becomes prohibitively large quickly (see below in Figure 17). The stability theorem holds by construction for the modules, and we can show that a matching exists between indecomposables, however a natural question is what is a natural distance between the indecomposables – the support is one candidate, however a more refined valuation may exist.

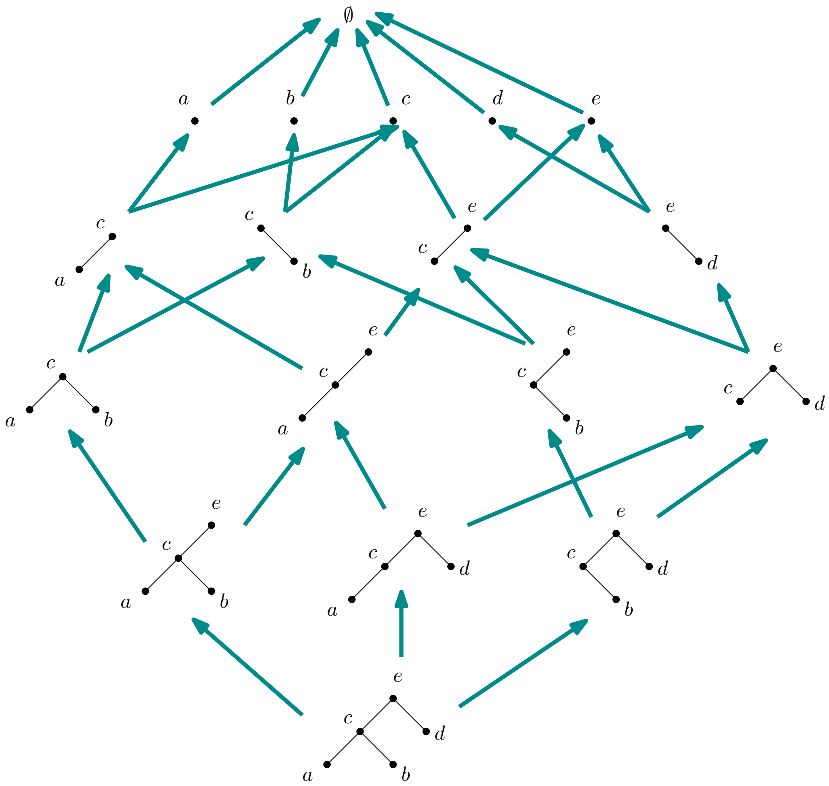

Figure 17 The Heyting algebra of a tree with 5 vertices. Note the arrows go “the wrong way” since at the sheaf-level we have contravariance so the arrows reverse direction..

Induced Barcodes & Towers of Matchings

In 2005, Cohen-Steiner, Edelsbrunner, and Harer introduced a stability result for the persistent homology of real-valued functions The result has become a centerpiece of the theory of persistent homology and topological data analysis, and has been the basis for much subsequent theoretical work on persistence. In subsequent work, it ahs been shown that stability is an immediate consequence of a purely algebraic result, which we will call the algebraic stability of persistence (ASP). The ASP asserts that if there exists a δ-interleaving (a sort of “approximate isomorphism”) between two persistence modules then there exists a δ-matching (approximate isomorphism) between their barcodes.

We have extending the work, establishing a further understanding of the relationships between persistence diagrams. The centrepiece of this work is the so-called **Isometry Theorem.** This is the fact, the converse to the ASP also holds: There exists a δ-interleaving between two persistence modules if and only if there exists a δ-matching between their barcodes. The ASP and its converse are together known as the isometry theorem. A slightly weaker formulation of the isometry theorem establishes a relationship between the interleaving distance (a pseudometric on persistence modules) and the bottleneck distance (a widely studied pseudometric on barcodes): It says that the interleaving distance between any two persistence modules is equal to the bottleneck distance between their barcodes.

The isometry theorem is interesting in part because the definition of the interleaving distance extends to a variety of generalized persistence settings where the direct definition of the bottleneck distance does not. For example, interleaving distances can be defined on multidimensional persistence modules and filtered topological spaces. The work is also connected to the **tower of matchings,** which was used in the work on the persistence of a self-map.

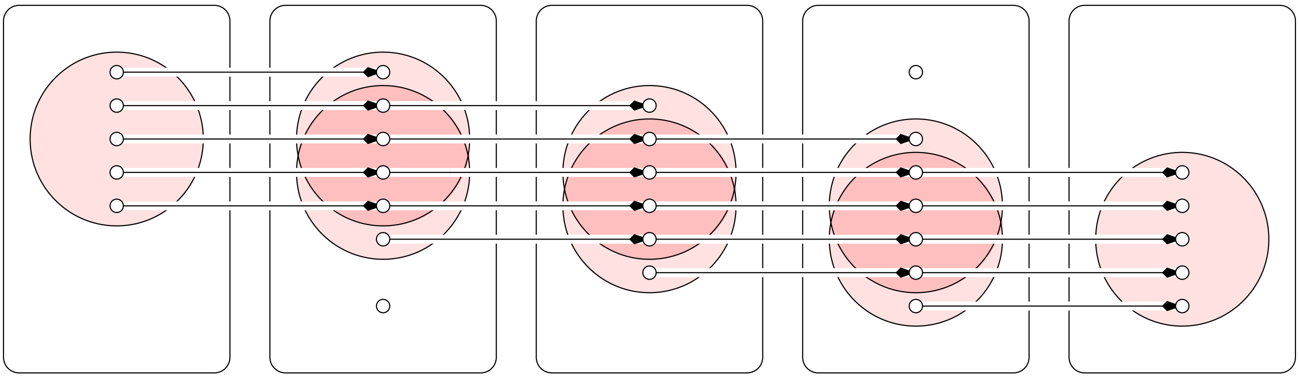

Figure 18 A barcode or persistence diagram can be seen as a matching of bases across different vector spaces. This viewpoint yields several interesting results including the non-functoriality of the category of persistence diagrams.

Semi-supervised learning

We developed a framework for clustering that strongly simplifies the transformation of an unsupervised clustering scheme to a semi-supervised scheme using the algebraic presentations to tag a subset of known cluster membership labels as assistance for both parameter selection and a topological identification step to produce a guided clustering from the provided labels.

Machine learning tasks can often be divided into three class: supervised, unsupervised and semi-supervised. Supervised learning is perhaps the most common: based on some examples, an algorithm learns and adapts to the task at hand. Typical examples include classification and estimation. Approaches usually involve solving certain optimization problems which may minimize error, bias or some other loss function. Unsupervised learning, on the hand, is more closely associated to exploratory data analysis, where we search for the underlying structure of the data. The most prominent examples are dimensionality reduction techniques, both linear (PCA) and non-linear (ISOMAP, LLE).

Clustering can be thought of as a coarse non-linear dimensionality reduction technique. It divides the data into discrete categories, which hopefully have some intrinsic meaning. It is perhaps one of the most studied problems in machine learning and many different approaches and definitions have been proposed. Semi-supervised learning falls between the two extremes. It describes a common setting today: we are given a large amount of data with a small labelled subset. The idea behind semi-supervised learning is that we should be able to do better than either just performing supervised learning on labelled data and less labour intensive than performing exploratory data analysis on the unlabelled data. We developed an algebraic framework based on ideas from applied topology to model semi-supervised learning. In particular, we concentrate on the problem of semi- supervised clustering and show how labelled data can be used to construct clusterings which are consistent both with the underlying structure of space of the data we are working with as well as the labels of the data.

Local Persistent Homology

Another area of machine learning we investigated was stratification learning (or mixed manifold learning). In this setting, a point cloud is assumed to be sampled from a mixture of (possibly intersecting) manifolds. The objective is to recover the different pieces, often treated as clusters, of the data associated with different manifolds of varying dimensions. An important tool in for this is **local homology.**

The local homology groups at a point are defined as the relative homology groups of a neighbourhood modulo its boundary. The local homology is a direct limit of relative homology groups as the neighbourhood size goes to zero. To make this more computational, we adapted two multi-scale notions of this concept based on persistence, which we referred to as the r-filtration and the α-filtration. Importantly, we derived sampling conditions that are appropriate to compute the persistence diagrams with respect to these filtrations, therefore approximating the local homology.

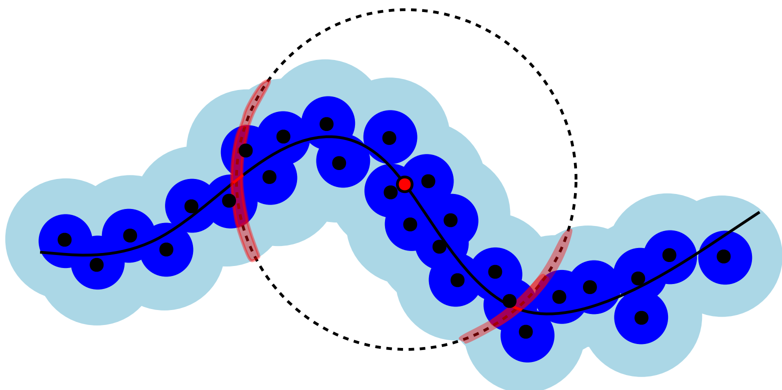

Figure 19 The α-filtration around the red point. By using various equivalences, we can avoid having to estimate the boundary (shaded red region) directly.

This work also allows us to use local persistence in a sheaf theoretic setting which is an important step towards stratification learning.

Morse-Conley-Forman Theory

Forman introduced the concept of a combinatorial vector field on a CW complex, introduced the concept of a chain recurrent set and proved Conley type generalization of Morse inequalities for basic sets consisting of critical cells and periodic trajectories. Conley theory is a generalization of Morse theory to the setting of non-necessarily gradient or gradient-like flows on locally compact metric spaces. In this theory, the concept of a critical point is replaced by a more general concept of an isolated invariant set and the Morse index of a critical point by the Conley index of an isolated invariant set.

Combinatorial vector fields seem to be a natural tool for a concise approximation and description of the dynamics of differential equations and more generally flows. We have developed the concept of an isolated invariant set and the Conley index in the case of a combinatorial vector field on the collection of simplexes of a simplicial complex. The aim of the work is to combine the ideas of Forman with some classical concepts of topological dynamics in order to obtain a tool for combinatorization of classical dynamics. In particular, in order to overcome the mentioned limitations of combinatorial vector fields in the approximation of differential equations we introduce combinatorial multivector fields, a generalization of Forman's combinatorial vector fields. Furthermore, we define attractors, repellers, Morse decompositions and present Morse inequalities for such Morse decompositions. These ideas are novel not only for combinatorial multivector fields but even for combinatorial vector fields.

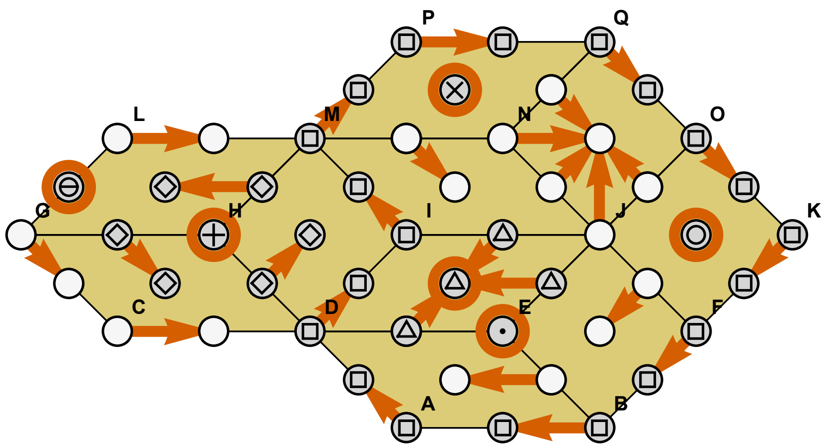

Figure 20 A combinatorial vector field on a CW complex. This may have attractors, repellers and even periodic orbits.

Applications and Verification

We briefly highlight several applications, mostly in the area of robotics but also provide a summary of results in other areas.

Grasping

We applied the topological tools to the problem of simultaneous clasp and motion planning on unknown objects with holes. Clasping an object enables a rich set of activities such as dragging, toting, pulling and hauling which can be applied to both soft and rigid objects. To this end, we defined a virtual linking measure, which characterizes the spatial relation between the robot hand and object. The measure utilizes a set of closed curves arising from an approximately shortest basis of the object's first homology group. We define task spaces to perform collision-free motion planning with respect to multiple prioritized objectives using a sampling-based planning method. We demonstrate that it is possible to equip robots with a richer repertoire of grasps for object manipulation and interaction with the environment.

Example hand representations and finger curves are depicted in Figure 21. The five real-sized experiment objects are displayed in Fig. 22 along with their representation created from 2000 points sampled on the object's visible surface. From the detected shortest loops, only the loop marked in red is used for clasping.

The results illustrate how a topological abstraction, in this case, that of a loop, allows for us to understand something important about the object in question – in this case, where it can be grasped.

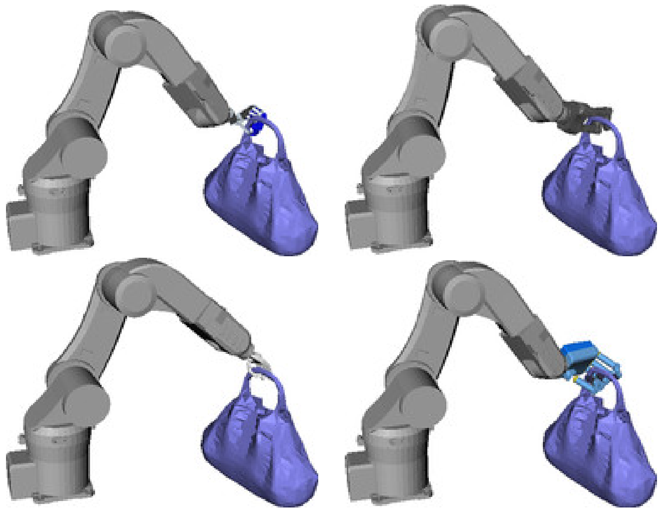

Figure 21 A few examples of our approach for clasp planning with various robotic hands. Here, the topological feature describes the purse's handle.

Figure 22 Left to right: real size objects used for experiments and the number of detected loops: Purse (1), Bag (1), Travel Bag (1), Lawn Mower (2), Chair (7). Bottom row: collision model and selected loops used as object representation in the experiments.

Gait Analysis using Cohomology

We applied the techniques of WP1 to detect, parameterize and interpolate periodic motion patterns obtained from a motion capture sequence. Using our developed framework, periodic motions such as walking and running gaits or any motion sequence with periodic structure such as cleaning, dancing etc. can be detected automatically and without manual marking of the period start and end points. Our approach constructs an intrinsic parameterization of the motion and is computationally fast. Using this parameterization, we are able to generate prototypical periodic motions, interpolate between various motions, yielding a rich class of "mixed" periodic actions and construct signatures of motions

Our approach is based on a persistent cohomology based method, which enables us to recover circular coordinates of motions. Our contributions are also outlined in a video summary, accessible at http://www.youtube.com/watch?v=NGQ-M2gdibQ. We process motions using the pipeline outlined in Figure 23. The input to our approach is given by a sequence of points describing a piecewise linear (PL) motion trajectory and our approach can be split into three main stages: preprocessing, parameterization and output.

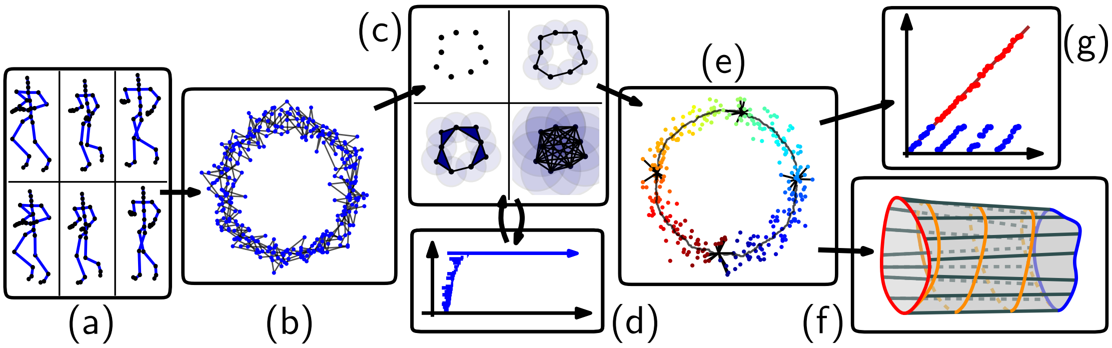

Figure 23 We display our topological processing pipeline for periodic motions: The input motion capture data (a) with a near-periodic trajectory (b) in a high-dimensional configuration space is visualized (in 2D) in the first two steps. We then construct a Vietoris-Rips filtration (c) and compute its persistent cohomology (d). We then produce circle-valued coordinate functions (e) and use these functions for several applications including describing the motion itself (f) and its periodicity, as well as to automatically align motions for blending or interpolation (g).

In addition, to operations such as averaging, blending and interpolation (illustrated in Figure 24), the coordinates can also be used to segment motions where there are several submotions, by tracking their values over time. Below we show two examples, a motion consisting of walking, running, and hopping and a motion consisting of different punches – where we found an uppercut and a jab.

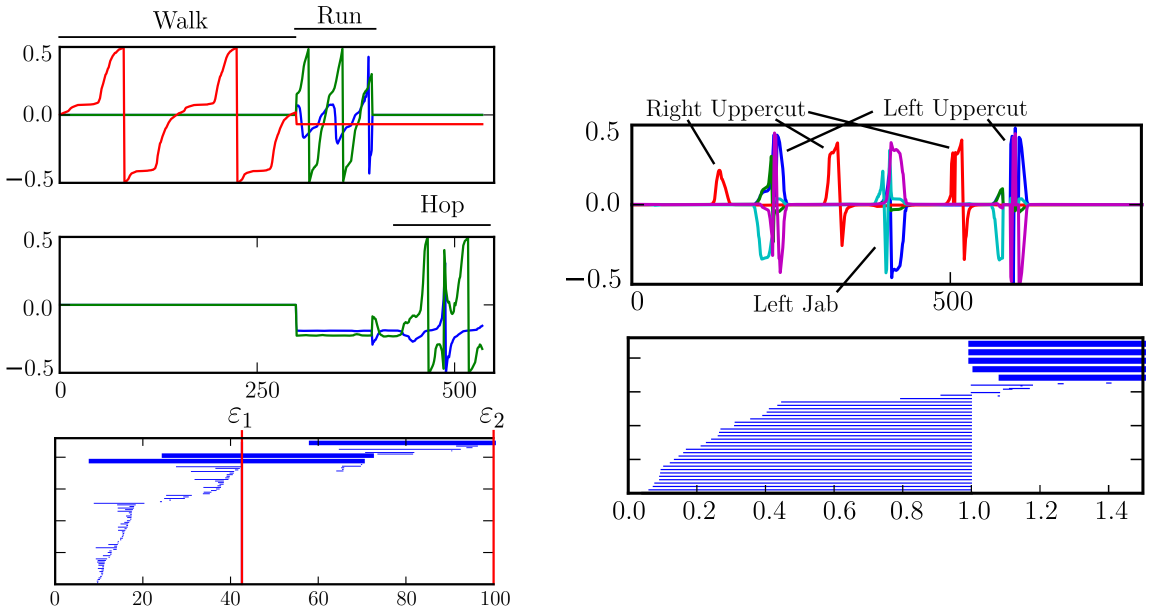

Figure 24 (Left) We combined motions a regular walk, jogging, and hopping (regular walk), into a combined sequence and computed coordinates for the entire sequence. The smaller scale captured the slower walking and running motions well, but missed the faster jumping motion at the end, while the larger scale missed the slow motions and detects the fast motion instead. (Right) A boxer throwing several different types of punches. The various circular coordinate excitations can differentiate between the two types of left punches also – even though this relationship is not a straightforward correspondence.

Multiscale Trajectory Classification

Another application we studied where the equivalence classes of trajectories in a configuration space. The study of these classes has recently received attention in robotics since they allow a robot to reason about trajectories at a high level of abstraction. While recent work has approached the problem of topological motion planning under the assumption that the configuration space and obstacles within it are explicitly described in a noise-free manner, we focus on trajectory classification and present a sampling-based approach which can handle noise, which is applicable to general configuration spaces and which relies only on the availability of collision free samples. Unlike previous sampling-based approaches in robotics which use graphs to capture information about the path-connectedness of a configuration space, we construct a multi-scale approximation of neighborhoods of the collision free configurations based on filtrations of simplicial complexes.

Our approach thereby extracts additional homological information, which is essential for a topological trajectory classification. By computing a basis for the first persistent homology groups, we obtain a multi-scale classification algorithm for trajectories in configuration spaces of arbitrary dimension. We furthermore show how an augmented filtration of simplicial complexes based on a cost function can be defined to incorporate additional constraints. We tested and evaluated of our approach in 2, 3, 4 and 6 dimensional configuration spaces simulation and using a Baxter robot. Some results are shown below in Figure 25.

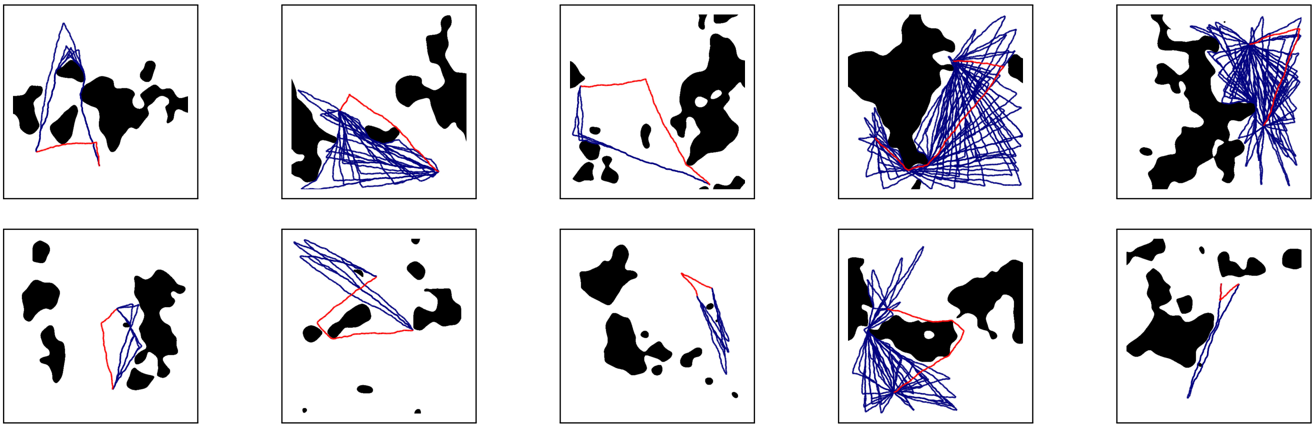

Figure 25 We display example worlds and examples of paths, which were determined to lie in a single class (in blue) at a particular filtration value. The space was constructed from 100000 samples and the classes are computed using the indicated red trajectories.

This work illustrates how the equivalences naturally considered in topological invariants can serve to encode high-level information about a complex problem.

Development of Algorithms

A key point in our work was to ensure that we could compute persistence on datasets with large input sizes (for applications). We have a number of algorithmic and implementation advances to report as well as a better theoretical understanding of the problem of computing persistence.

Our contribution is a novel algorithm that incorporates two optimization techniques, **Chunk** and **Clear and Compress**, and is also suitable for parallelization (on a shared memory machine). This has been shown to greatly speed up computing persistence. Both are publicly available at http://phat.googlecode.com/ - it has since been moved to BitBucket, pending the next release.

We also developed a scalable algorithm for computing persistent homology in parallel in a distributed memory environment. This method the computation of much larger instances than using existing state-of-the art algorithms on a single machine, by using sufficiently many computing nodes such that the data fits into the distributed memory. While overcoming the memory bottleneck is the primary purpose of our approach, we aim for a time-efficient solution at the same time. This was implemented in the library **DIPHA**, which is publicly available at https://code.google.com/p/dipha/. The hosting has moved to BitBucket and will be publicly released there shortly.

Finally, we provide an analysis of the persistence algorithm in the special case of complexes, which have small separators. This relies on a technique known as generalized nested dissection which was developed as a technique to reduce the effect of fill-in in Gaussian elimination and related matrix reduction problems for instances which have certain structure. We show for the first time that there exists a deterministic algorithm, which beats matrix multiplication time.

Other Applications

We also investigated other applications, including but not limited to using persistence diagrams as a feature for epidemiological data (shown below in Figure 26), using persistence diagrams to look at genetic data, and finally doing topological data analysis using Bregman diverences which connects applied topology with information geometry and promises more refined information to be extracted for textual data (the original motivation for this investigation).

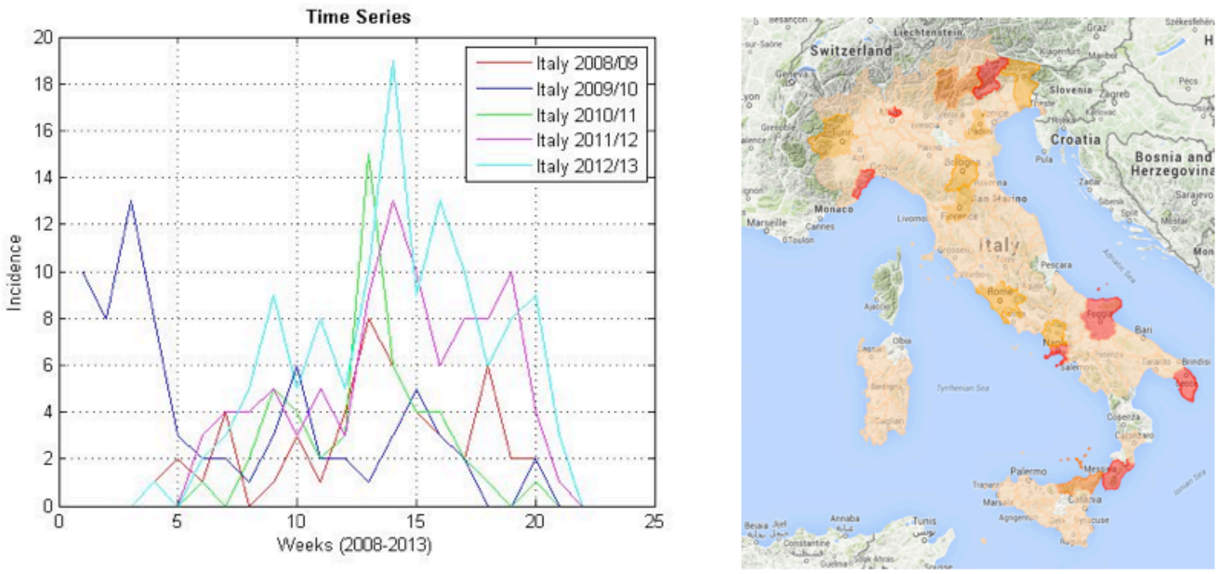

Figure 26 Using Takens' embedding and persistence as a feature for quantifying similarity between flu seasons.

Other applications include but are not limited to genomics data and materials science.

Societal Impact and socio-economic

The project has made an impact on several fields and the results have been in a number of venues over the time of the project. There have been numerous papers published and many more submitted or to appear. The consortium as a whole has organized several summer schools, tutorials and workshops, including a satellite at the European Conference on Complex Systems. Other workshops were organized in related areas including computer science and applied mathematics.

In terms of exploitation, there were inquiries into patenting the gait analysis. Unfortunately, given the early stage of development and costs involved, this avenue did not prove feasible. The concentration on foundational work implies that such exploitation will occur after the project.

There were several all-hands meetings for TOPOSYS

- TOPOSYS kick-off meeting (Ljubljana)
- TOPOSYS first year meeting (Technion)
- TOPOSYS second year meeting (Vienna)
- TOPOSYS application meeting (Stockholm)
- TOPOSYS third year meeting (Krakow)

The main role of the project kick off meeting was for the consortium to get to know each other. To this end, all the partners presented tutorial style presentations of their previous work, and how they envision their role in the project. It is worth noting that the **kick-off meeting covered by the Slovenian Press Agency.** The meetings were mostly closed, with the exception of the applications meetings which had several other researchers from KTH present.

While the number of all-hands meetings was low, there were numerous meetings, however mostly done as focused visits or within the context of some other event. It was found this was more effective.

Tutorials and Workshops

We first give an overview of the tutorials and workshops where there was a substantial presence by the project

Topological Data Analysis - The 30th International Conference on Machine Learning (ICML 2013, Atlanta, USA) – This tutorial provided an introduction to Topological Data Analysis (TDA) to the machine learning community. The idea behind TDA is to extract robust topological features from data and use these summaries for modelling the data. Topological features can quantify qualitative changes in data and formalize its global structure. We covered one of the main tools in TDA, persistent homology, including the necessary background. In particular, no prior knowledge of algebraic topology was required or assumed. The focus was be on how these methods relate to and may be applied in machine learning. The tutorial also gave a brief overview of some of the available tools and current relevant directions of research in the field.

Introduction to Statistics and Probability for Topologists (IMA) –Minneapolis, Minnesota - Applying topological methods for the analysis of complex data sets require two skill sets: a knowledge of topology and an understanding of the basically stochastic nature of data, which is typically sampled from some underlying random structure. Acknowledging the fact that topologists are only rarely trained in statistical methodology, the aim of this tutorial will be to acquaint topologists with the basic concepts and tools of probability and statistics, as well of some of the more advanced techniques of specific interest in topological data analysis.

Topics to be covered will include basic probability, statistical inference (both classical frequentist and Bayesian approaches), and an introduction to Gaussian and Markov stochastic processes along with Markov chain Monte Carlo as a simulation and inference tool. Among the more advanced topics to be described will be handling statistical outliers, graphical models, statistical clustering tools, and spatial dependence.

Algebraic Topology- Methods, Computation and Science 6 (ATMCS6) - 26.5.2014 – 30.5.2014 – Applied and computational topology refers to the adaptation of topological ideas and techniques to study problems in science and engineering. A particular focus is on using invariants and methods of algebraic topology to understand large high-dimensional data sets. The further development of topological techniques for use in applications and the creation of new areas of application in the subject are amongst the goals of this workshop.

The workshop will bring together leading researchers in this emerging discipline as well as providing an opportunity for young mathematicians to get involved in it. In past years, the ATMCS conference has been very successful in providing a forum for cutting-edge research to be disseminated; attendance tends to represent a broad swath of the diverse research community which works in this area.

Topological Methods for Machine Learning - The 31th International Conference on Machine Learning (ICML 2014, Beijing, USA) - Computational topology saw three major developments in recent years: persistent homology, Euler calculus and Hodge theory. Persistent homology extracts stable homology groups against noise; Euler Calculus encodes integral geometry and is easier to compute than persistent homology or Betti numbers; Hodge theory connects geometry to topology via optimization and spectral method. All three techniques are related to Morse theory, which is inspiring new computational tools or algorithms for data analysis. Computational topology has inspired a number of applications in the last few years, including game theory, graphics, image processing, multimedia, neuroscience, numerical PDE, ranking, robotics, voting theory, sensor networks, and natural language processing.

Workshop on Computational Topology and Data Analysis (CG-Week: Symposium of Computational Geometry 2013) – Computational topology has played a synergistic role in bringing together research work from computational geometry, algebraic topology, data analysis, and many other related scientific areas. In recent years, the field has undergone particular growth in the area of data analysis. The application of topological techniques to traditional data analysis, which before has mostly developed on a statistical setting, has opened up new opportunities. This workshop is intended to cover this aspect of computational topology along with the developments of generic techniques for various topology-centered problems.

Members of group at IST were among the organizers of the workshop which was part of the CG Week at the Symposium of Computational Geometry, which is the top theoretical Computer Science conference for geometry and topology.

SIAM Symposium on Applied Algebriac Geometry (AAG13) special session on Applied and Computational Topology, August 2013 (Fort Collins, USA) – Applied and computational topology is a vibrant research area that's gained momentum over the last decade. A core aim is data analysis by way of understanding the shape of the data. To devise robust techniques, researchers are interested in questions of stability of topological descriptors. To process modern datasets, the field is interested in efficient algorithms. The goal of the mini- symposium is to create a forum for young researchers to present recent developments in the field.

IMA Topological Data Analysis - October 7-11, 2013 - Recently, several techniques have emerged to try to deduce algebraic and/or geometric properties of a space from finite metric subspaces of sampled points. For example, starting with a manifold, a basic problem is to compute approximations to its homology, or local coordinate charts. A central issue in this area is the dependence of the invariants computed on noise or on the sampling procedure. This has led to the development of new algebraic topological invariants (e.g., persistent homology), and to the introduction of statistical ideas and techniques into topological questions. From the perspective of basic research, directions in this area include the study of the homotopy theory of point cloud data directly, as well as understanding the statistics associated to sampling, sampling with noise, and the relationship between the sampling procedure and the underlying geometry of the space in question. These ideas and related methods have led to a series of revealing analyses of experimental data (e.g., from neuroscience, fluid dynamics, genomics, and materials science). The interaction with actual data also brings to the forefront questions about computational complexity, large data sets, and computational procedures for handling noise.

The intended audience of the workshop is mathematicians and statisticians, as well as researchers working in application areas such as those listed above. The goals of the workshop include advancing the merger of statistical and topological techniques, extending the range and depth of applications, and expanding awareness of the challenges and new directions of the field.

IMA Algebraic Topology in Dynamics, Differential Equations, and Experimental Data February 10-14, 2014 - Homological quantities provide robust computable invariants of dynamical systems well-adapted to numerical methods. As a consequence, several groups have actively implemented algebraic topological invariants to characterize the qualitative behavior of dynamical systems. Examples include tracking patterns of nodal domains, proving the existence of invariant sets in infinite-dimensional systems, and generating forcing theories for invariant sets based on topological indices. Extensions of Morse-theoretic ideas due to Conley (in the finite-dimensional setting) and Floer (in the infinite-dimensional cases) are especially relevant and are active foci of current techniques in applied dynamical systems. This workshop will bring together researchers in topological methods with those working in differential equations and the physical sciences.

SAMSI LDHD: Topological Data Analysis: February 3-7, 2014 – Topological Data Analysis began a little more than 20 years ago with the introduction of persistence by Edelsbrunner, Lestcher and Zomorodian. Since that time the field has grown significantly. Many strong theorems, numerous algorithms, and pieces of software have been created in that period. More recently, the field has considered stochastic modeling and analysis, which has resulted in interactions with Statistics, Probability, and Computer Science. This workshop focused on the interaction of Statistics and Probability with this new area of Applied Mathematics. The goals were to familiarize people in both areas with techniques from the other and to understand what future directions are the most promising.

Topics included:

- Statistics and persistent homology
- Topological data analysis and inference in dynamical systems and time series
- Topological data analysis and shape statistics
- Topological data analysis and network models
- Probability models for random topology objects
- Statistical applications of Hodge theory
-

IMA Topological Systems: Communication, Sensing, and Actuation March 3-7, 2014- Continuing innovations in integrated circuit miniaturization and low-power local networking have made feasible the vision of distributed computers consisting of numerous cheap processing elements with limited local connectivity. Motivating practical examples of such systems include environmental sensor networks, cell phones, and multi-agent robotics. One pressing problem is how to perform system-wide analysis, assembling local information to infer or actuate global results. Recent advances have leveraged local-to-global principles from algebraic topology to provide robust algorithms for engineering systems. These include 1) homological methods for sensor networks, 2) topological complexity for robot motion planning, 3) simplicial complexes for planning in the presence of uncertainty, and 4) sheaf theory for network data aggregation and optimization.

This workshop will focus attention on the best and newest topological tools for applications in robotics, sensor networks, communications networks, signal processing, and more. The workshop will provide an excellent opportunity for curious researchers from diverse backgrounds to learn new mathematics or new applications. The goal of the workshop is to exchange ideas, problems, techniques, and motivations.

Dynamics, topology and computations – (Bedlewo, Poland June 15-20) - The conference is devoted to computational aspects of dynamics and topology. TOPOSYS was represented by four out of five partners. This conference covers computational aspects of both dynamical systems and applied topology.

The 4th Annual Minisymposium on Computational Topology - Sheaves and Categories - Symposium of Computational Geometry (Eindhoven, Netherlands) - This mini-symposium concentrates on the application of topological techniques to traditional data analysis. This mini-symposium concentrates on how computational geometry and topology fertilize each other, by covering recent advances on the interface of both areas. In this area, TOPOSYS presented on sheaves and categories, specifically introducing the notion of sheaves and related algorithms as well the work on topoidal persistence.

XXI Oporto Meeting on Geometry, Topology and Physics Foundations of Computational Mathematics (Oporto, Portugal) – This is a meeting, which brings together researchers from mathematicians and physicists.

Discrete, Computational and Algebraic Topology (Copenhagen Denmark) – TOPOSYS had a presence in terms of invited speakers from the Jozef Stefan Institute and Jagiellonian University. The work on the topoidal foundations of persistence as well as Forman-Morse-Conley theory were presented.

ECCS

We highlight the contributions at the European Conference of Complexity Science.

Topological Methods in Complexity Science (TMCS 2013) - Satellite of the European Conference on Complexity Science – Many complex systems are characterized by multi-level properties that make the study of their dynamics and of their emerging phenomena a daunting task. The huge amount of data available in modern sciences is expected to promote rapid progress in these areas, even though the nature of the data varies. Given the heterogeneity of the data, topological features often clearly convey important qualitative features of a system, while retaining the quantitative rigour.

This workshop aims at offering an up-to-date view on the study of complex multi-level systems via topological methods such as persistent (co)homology and techniques from topological dynamics.

This was co-organized with Francesco Vaccarino (ISI) who is part of the TOPDRIM project, which is also part of the DYM-CS.

TOPOSYS also had a presence on the panel discussion during the DYM-CS special session at ECCS both in 2013 and 2014. In 2015, the location of the USA proved not feasible. We believe we have made a good step in establishing a presence for computational topology in the European Complexity Community as well as informing the computational topology community about Complexity Science.

Summer Schools

Summer School on Computational Topology and Topological Data Analysis (Ljubljana, Slovenia, July 1st - July 4th 2013) - The main goal of the Summer school was to introduce the basics of new and attractive research fields in Computational Topology and Topological Data Analysis to students and young researchers in Mathematics and Computer Science. The school was aimed mainly at Master or early PhD level students.

The Summer school consisted of a series of lectures by invited lecturers discussing various topics in Computational Topology, followed by afternoon lab work, where the students were given the opportunity to work together on assigned projects related to the topics of the lectures. The general orientation of the summer school was towards applications of the methods of Computational

Topology, like persistent homology, discrete Morse theory and topological robotics to the solution of specific problems from different areas, like Image Analysis and Bioinformatics, as well as on specific algorithms.

Summer School on Computational Topology June 22-25, 2015 Faculty of Computer and Information Science Ljubljana – Slovenia - This was the second summer school organized in Ljubljana The main goal of the Summer school was to introduce the basics of new and attractive research fields in Computational Topology and Topological Data Analysis to students and young researchers in Mathematics and Computer Science. The school was aimed mainly at Master or early PhD level students.

The Summer school consisted of a series of lectures by invited lecturers discussing various topics in Computational and Applied Topology. This year there was a greater emphasis on dynamics. Francesco Vaccarino (ISI) who is part of the TOPDRIM project, which is also part of the DYM-CS.

Cooperation with the IMA and ACAT

The IMA(Institute of Mathematics and Applications, Minneapolis, USA) had a thematic year on Applied Topology during the course of the project. The first tutorial and workshop took place at the beginning of October 2013, with a very large attendance (over 150 people), indicating the intense interest in the field. Over the course of the year, there will be numerous interactions at the institute as well as the numerous workshops throughout the year – many with topics of direct relevance to complexity science.

Over the course of the year, there were a number of workshops throughout the year – many with topics of direct relevance to complexity science. Many of the dissemination activities concentrated on the thematic year and the related program at SAMSI (Statistical and Applied Mathematical Sciences Institute), which focused more on statistics – the Program on Low-dimensional Structure in High-dimensional Systems (LDHD)

There was cooperation between the ACAT program in organizing a summer schools in Ljubljana at which numerous TOPOSYS members participated in. There were several other ACAT sponsored events which where well attended by the TOPOSYS project.

Beyond the Project

Applied and computational topology has continued to have a growing presence in a variety of fields including complexity science. The number of workshops dedicated to its study and its applications continues to grow and the results produced in this project have had and will continue to have an impact on the field as well as on its application fields. For example, we see that there are tight connections with material science. In the near term, we believe that the statistical and probabilistic results will have an impact in our understanding of networks resulting in a better models and an understanding of phase transitions as a topological phenomena.

The collaborations established during the project will continue and be extended, with several TOPOSYS alumni becoming faculty. This is also true for collaboration with other projects – with several new longer term collaborations being established. Exploitation beyond available tools will be seen as we investigate how new visualization technology may be leveraged. The improvement in scalability makes this possible for the first time. These avenues will continue to be explored.

The field is continuing to grow and we believe that new connections and applications will be found.

Code

Finally, we have made several code bases available. The two most notable examples are PHAT and DIPHA for computing persistence available at

- https://code.google.com/p/phat/
- https://dipha.googlecode.com/

Note that these are being migrated to Bitbucket. While available, the code for the self-map and circular coordinates is being improved for scalability and usabililty before full release (it is still available upon request). Development of both is continuing including the improvement of usability such as availability of complex constructions and relative persistence.

Website

Address of the website: www.toposys.org

Logo

TOPOSYS

4.2 Use and dissemination of foreground

A plan for use and dissemination of foreground (including socio-economic impact and target groups for the results of the research) shall be established at the end of the project. It should, where appropriate, be an update of the initial plan in Annex I for use and dissemination of foreground and be consistent with the report on societal implications on the use and dissemination of foreground (section 4.3 – H).

The plan should consist of:

- Section A

This section should describe the dissemination measures, including any scientific publications relating to foreground. **Its content will be made available in the public domain** thus demonstrating the added-value and positive impact of the project on the European Union.

- Section B

This section should specify the exploitable foreground and provide the plans for exploitation. All these data can be public or confidential; the report must clearly mark non-publishable (confidential) parts that will be treated as such by the Commission. Information under Section B that is not marked as confidential **will be made available in the public domain** thus demonstrating the added-value and positive impact of the project on the European Union.