Perspectives in Persistent Homology

Bastian Rieck

Looking back

Persistent homology

The first decade

The very beginning


Topological Persistence and Simplification

H. Edelsbrunner et al., ‘Topological persistence and simplification’

Persistent homology

The first decade

Computing persistent homology in all dimensions

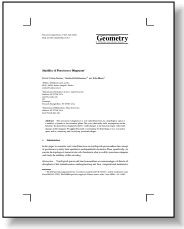
A. J. Zomorodian and G. Carlsson, ‘Computing persistent homology’

2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012

Persistent homology

The first decade

Hausdorff stability of persistence diagrams



D. Cohen-Steiner et al., ‘Stability of persistence diagrams’

2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012

Persistent homology

The first decade

Mapper algorithm

Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition

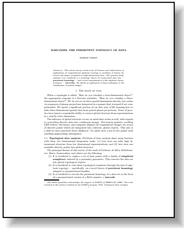
G. Singh et al., ‘Topological methods for the analysis of high dimensional data sets and 3D object recognition’

2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012

Persistent homology

The first decade

Barcodes


BARCODES: THE PERSISTENT TOPOLOGY OF DATA

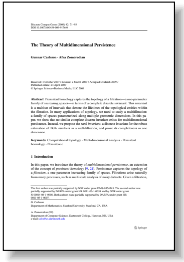
R. Ghrist, ‘Barcodes: The persistent topology of data’

2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012

Persistent homology

The first decade

Multidimensional persistence


Discrete Comput Geom (2009) 42: 71–93
DOI 10.1007/s00454-009-9176-0

The Theory of Multidimensional Persistence

Gunnar Carlsson · Afra Zomorodian

Received: 1 October 2007 / Revised: 2 March 2009 / Accepted: 2 March 2009 /
Published online: 24 April 2009

Abstract Persistent homology captures the topology of a filtration—a one-parameter family of increasing spaces—in terms of a complete discrete invariant. This invariant is a multiset of intervals that denote the lifetimes of the topological entities within the filtration. In many applications of topology, we need to study a multifiltration: a family of spaces parameterized along multiple geometric dimensions. In this paper, we show that no similar complete discrete invariant exists for multidimensional persistence. Instead, we propose the *rank invariant*, a discrete invariant for the robust estimation of Betti numbers in a multifiltration, and prove its completeness in one dimension.

Keywords Computational topology · Multidimensional analysis · Persistent homology · Persistence

1 Introduction

In this paper, we introduce the theory of *multidimensional persistence*, an extension of the concept of *persistent homology* [9, 21]. Persistence captures the topology of a *filtration*, a one-parameter increasing family of spaces. Filtrations arise naturally from many processes, such as multiscale analysis of noisy datasets. Given a filtration,

The first author was partially supported by NSF under grant DMS-0354543. The second author was partially supported by DARPA under grant HR 0011-06-1-0038 and by ONR under grant N 00014-08-1-0908. Both authors were partially supported by DARPA under grant HR 0011-05-1-0007.

G. Carlsson
Department of Mathematics, Stanford University, Stanford, CA, USA

A. Zomorodian (✉)
Department of Computer Science, Dartmouth College, Hanover, NH, USA
e-mail: afra@cs.dartmouth.edu

Springer

G. Carlsson and A. J. Zomorodian, ‘The theory of multidimensional persistence’

2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012

Persistent homology

The first decade

Extended persistence


Found Comput Math (2009) 9: 79–103
DOI 10.1007/s10208-008-9027-z

FOUNDATIONS OF COMPUTATIONAL MATHEMATICS

Extending Persistence Using Poincaré and Lefschetz Duality

David Cohen-Steiner · Herbert Edelsbrunner · John Harer

Received: 3 October 2006 / Revised: 22 January 2008 / Accepted: 1 February 2008 /
Published online: 4 April 2008

Abstract Persistent homology has proven to be a useful tool in a variety of contexts, including the recognition and measurement of shape characteristics of surfaces in $\mathbb{R}^3$. Persistence pairs homology classes that are born and die in a filtration of a topological space, but does not pair its actual homology classes. For the sublevelset filtration of a surface in $\mathbb{R}^3$, persistence has been extended to a pairing of essential classes using Reeb graphs. In this paper, we give an algebraic formulation that extends persistence to essential homology for any filtered space, present an algorithm to calculate it, and describe how it aids our ability to recognize shape features for codimension 1 submanifolds of Euclidean space. The extension derives from Poincaré duality but generalizes to nonmanifold spaces. We prove stability for general triangulated spaces and duality as well as symmetry for triangulated manifolds.

Communicated by Konstantin Mischaikow.

Research by all three authors is partially supported by DARPA under grant HR0011-05-1-0007.
Research by the second author is also partially supported by NSF under grant CCR-00-86013.

D. Cohen-Steiner
INRIA, 2004 Route des Lucioles, BP93, Sophia-Antipolis, France

H. Edelsbrunner (✉)
Department of Computer Science, Duke University, Durham, USA
e-mail: edels@cs.duke.edu

H. Edelsbrunner
Geomagic, Research Triangle Park, NC, USA

J. Harer
Department of Mathematics and Center for Computational Science, Engineering and Medicine, Duke University, Durham, NC, USA

Springer

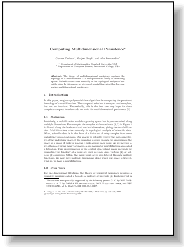
D. Cohen-Steiner et al., ‘Extending persistence using Poincaré and Lefschetz duality’

2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012

Persistent homology

The first decade

Algorithms for computing multidimensional persistence



Computing Multidimensional Persistence*

Gunnar Carlsson[1], Gurjeet Singh[1], and Afra Zomorodian[2]

[1] Department of Mathematics, Stanford University, USA
[2] Department of Computer Science, Dartmouth College, USA

Abstract. The theory of multidimensional persistence captures the topology of a multifiltration – a multiparameter family of increasing spaces. Multifiltrations arise naturally in the topological analysis of scientific data. In this paper, we give a polynomial time algorithm for computing multidimensional persistence.

1 Introduction

In this paper, we give a polynomial time algorithm for computing the persistent homology of a multifiltration. The computed solution is compact and complete, but not an invariant. Theoretically, this is the best one may hope for since complete compact invariants do not exist for multidimensional persistence [1].

1.1 Motivation

Intuitively, a multifiltration models a growing space that is parameterized along multiple dimensions. For example, the complex with coordinate (3, 2) in Figure 1 is filtered along the horizontal and vertical dimensions, giving rise to a bifiltration. Multifiltrations arise naturally in topological analysis of scientific data. Often, scientific data is in the form of a finite set of noisy samples from some underlying topological space. Our goal is to robustly recover the lost connectivity of the underlying space. If the sampling is dense enough, we approximate the space as a union of balls by placing ϵ-balls around each point. As we increase ϵ, we obtain a growing family of spaces, a one-parameter multifiltration also called a filtration. This approximation is the central idea behind many methods for computing the topology of a point set, such as Čech, *Rips-Vietoris* [2], or witness [3] complexes. Often, the input point set is also filtered through multiple functions. We now have multiple dimensions along which our space is filtered. That is, we have a multifiltration.

1.2 Prior Work

For one-dimensional filtrations, the theory of persistent homology provides a complete invariant called a *barcode*, a multiset of intervals [4]. Each interval in

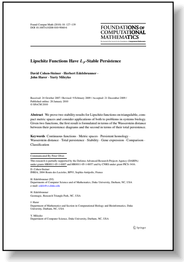
G. Carlsson et al., 'Computing multidimensional persistence'

2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012

Persistent homology

The first decade

Lipschitz stability of persistence diagrams


Found Comput Math (2010) 10: 127–139
DOI 10.1007/s10208-010-9060-6

FOUNDATIONS OF COMPUTATIONAL MATHEMATICS

Lipschitz Functions Have L_p-Stable Persistence

David Cohen-Steiner · Herbert Edelsbrunner · John Harer · Yuriy Mileyko

Received: 24 October 2007 / Revised: 9 February 2009 / Accepted: 21 December 2009 / Published online: 28 January 2010

Abstract We prove two stability results for Lipschitz functions on triangulable, compact metric spaces and consider applications of both to problems in systems biology. Given two functions, the first result is formulated in terms of the Wasserstein distance between their persistence diagrams and the second in terms of their total persistence.

Keywords Continuous functions · Metric spaces · Persistent homology · Wasserstein distance · Total persistence · Stability · Gene expression · Comparison · Classification

Communicated By Peter Olver.

This research is partially supported by the Defense Advanced Research Projects Agency (DARPA) under grants HR0011-05-1-0007 and HR0011-05-1-0057 and by CNRS under grant PICS-3416.

D. Cohen-Steiner
INRIA, 2004 Route des Lucioles, BP93, Sophia-Antipolis, France

H. Edelsbrunner (✉)
Departments of Computer Science and of Mathematics, Duke University, Durham, NC, USA
e-mail:

H. Edelsbrunner
Geomagic, Research Triangle Park, NC, USA

J. Harer
Department of Mathematics and Section in Computational Biology and Bioinformatics, Duke University, Durham, NC, USA

Y. Mileyko
Department of Computer Science, Duke University, Durham, NC, USA

Springer

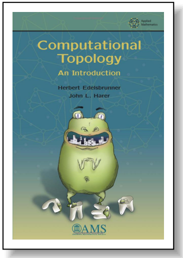
D. Cohen-Steiner et al., 'Lipschitz functions have L_p-stable persistence'

2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012

Persistent homology

The first decade

The book



H. Edelsbrunner and J. Harer, *Computational topology: An introduction*

2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012

Persistent homology

The first decade

Efficient Vietoris–Rips complex construction


Contents lists available at ScienceDirect

Computers & Graphics

Technical Section

Fast construction of the Vietoris-Rips complex

Afra Zomorodian

ARTICLE INFO

ABSTRACT

1. Introduction

A. J. Zomorodian, ‘Fast construction of the Vietoris–Rips complex’

Persistent homology

The first decade

`Mapper` for breast cancer analysis


PNAS

Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival

Monica Nicolau, Arnold J. Levine, and Gunnar Carlsson

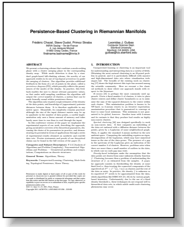
M. Nicolau et al., 'Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival'

2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012

Persistent homology

The first decade

Clustering meets persistent homology


Persistence-Based Clustering in Riemannian Manifolds

Frédéric Chazal, Steve Oudot, Primoz Skraba
INRIA Saclay – Île-de-France
4, rue Jacques Monod
91893 Orsay Cedex, France
firstname.lastname@inria.fr

Leonidas J. Guibas
Computer Science Dept.
Stanford University
Stanford, CA 94305, USA
guibas@cs.stanford.edu

ABSTRACT

1. INTRODUCTION

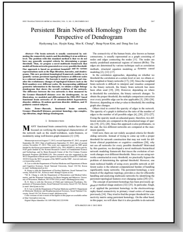
F. Chazal et al., 'Persistence-based clustering in Riemannian manifolds'

2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012

Persistent homology

The first decade

Connectivity analysis of the brain



Persistent Brain Network Homology From the Perspective of Dendrogram

Hyekyoung Lee, Hyejin Kang, Moo K. Chung*, Bung-Nyun Kim, and Dong Soo Lee

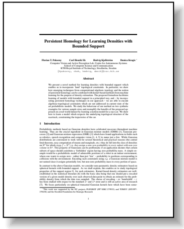
H. Lee et al., ‘Persistent brain network homology from the perspective of dendrogram’

2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012

Persistent homology

The first decade

Persistence-based bandwidth selection



Persistent Homology for Learning Densities with Bounded Support

Florian T. Pokorny Carl Henrik Ek Hedvig Kjellström Danica Kragic

Computer Vision and Active Perception Lab, Centre for Autonomous Systems
School of Computer Science and Communication
KTH Royal Institute of Technology, Stockholm, Sweden

Abstract

We present a novel method for learning densities with bounded support which enables us to incorporate 'hard' topological constraints. In particular, we show how emerging techniques from computational algebraic topology and the notion of persistent homology can be combined with kernel-based methods from machine learning for the purpose of density estimation. The proposed formalism facilitates learning of models with bounded support in a principled way, and – by incorporating persistent homology techniques in our approach – we are able to encode algebraic-topological constraints which are not addressed in current state of the art probabilistic models. We study the behaviour of our method on two synthetic examples for various sample sizes and exemplify the benefits of the proposed approach on a real-world dataset by learning a motion model for a race car. We show how to learn a model which respects the underlying topological structure of the racetrack, constraining the trajectories of the car.

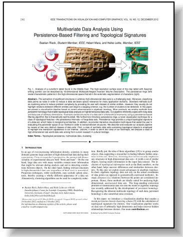
1 Introduction

F. T. Pokorny et al., 'Persistent homology for learning densities with bounded support'

Persistent homology

The first decade

Visual topological signatures


IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 18, NO. 12, DECEMBER 2012

Multivariate Data Analysis Using
Persistence-Based Filtering and Topological Signatures

Bastian Rieck, *Student Member, IEEE*, Hubert Mara, and Heike Leitte, *Member, IEEE*

B. Rieck et al., ‘Multivariate data analysis using persistence-based filtering and topological signatures’

2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012

Persistent homology

The second decade (almost)

`Mapper` applications


SCIENTIFIC REPORTS

OPEN

Extracting insights from the shape of complex data using topology

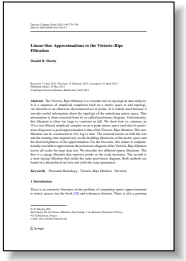
P. Y. Lum et al., ‘Extracting insights from the shape of complex data using topology’

2013 2014 2015 2016 2017 2018 2019

Persistent homology

The second decade (almost)

Sparse filtrations


Discrete Comput Geom (2013) 49:778–796
DOI 10.1007/s00454-013-9513-1

Linear-Size Approximations to the Vietoris–Rips Filtration

Donald R. Sheehy

Received: 13 July 2012 / Revised: 15 February 2013 / Accepted: 18 April 2013 /
Published online: 25 May 2013

Abstract The Vietoris–Rips filtration is a versatile tool in topological data analysis. It is a sequence of simplicial complexes built on a metric space to add topological structure to an otherwise disconnected set of points. It is widely used because it encodes useful information about the topology of the underlying metric space. This information is often extracted from its so-called persistence diagram. Unfortunately, this filtration is often too large to construct in full. We show how to construct an $O(n)$-size filtered simplicial complex on an n-point metric space such that its persistence diagram is a good approximation to that of the Vietoris–Rips filtration. This new filtration can be constructed in $O(n \log n)$ time. The constant factors in both the size and the running time depend only on the doubling dimension of the metric space and the desired tightness of the approximation. For the first time, this makes it computationally tractable to approximate the persistence diagram of the Vietoris–Rips filtration across all scales for large data sets. We describe two different sparse filtrations. The first is a zigzag filtration that removes points as the scale increases. The second is a (non-zigzag) filtration that yields the same persistence diagram. Both methods are based on a hierarchical net-tree and yield the same guarantees.

Keywords Persistent Homology · Vietoris–Rips filtration · Net-trees

1 Introduction

There is an extensive literature on the problem of computing sparse approximations to metric spaces (see the book [29] and references therein). There is also a growing

D. R. Sheehy (✉)
Inria Saclay Île-de-France, Bâtiment Alan Turing, 1 rue Honoré d'Estienne d'Orves,
91120 Palaiseau, France
e-mail: don.r.sheehy@gmail.com

Springer

D. R. Sheehy, 'Linear-size approximations to the Vietoris–Rips filtration'

Persistent homology

The second decade (almost)

Classifying lesions using barcodes


Computer Vision and Image Understanding

Classification of hepatic lesions using the matching metric

Aaron Adcock, Daniel Rubin, Gunnar Carlsson

A. Adcock et al., 'Classification of hepatic lesions using the matching metric'

2013 2014 2015 2016 2017 2018 2019

Persistent homology

The second decade (almost)

Persistence landscapes


Statistical Topological Data Analysis using Persistence Landscapes

Peter Bubenik

Department of Mathematics
Cleveland State University

Editor: David Dunson

Abstract

1. Introduction

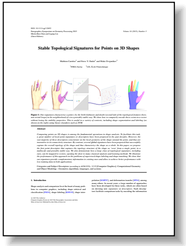
P. Bubenik, ‘Statistical topological data analysis using persistence landscapes’

2013 2014 2015 2016 2017 2018 2019

Persistent homology

The second decade (almost)

Persistence diagram vectorisation


Stable Topological Signatures for Points on 3D Shapes

M. Carrière et al., ‘Stable topological signatures for points on 3D shapes’

2013 2014 2015 2016 2017 2018 2019

Persistent homology

The second decade (almost)

Persistence-based metric for dimensionality reduction

Persistent Homology for the Evaluation of Dimensionality Reduction Schemes

B. Rieck and H. Leitte

B. Rieck and H. Leitte, ‘Persistent homology for the evaluation of dimensionality reduction schemes’

2013 2014 2015 2016 2017 2018 2019

Persistent homology

The second decade (almost)

Persistence scale space kernel


A Stable Multi-Scale Kernel for Topological Machine Learning

Jan Reininghaus, Stefan Huber
IST Austria

Ulrich Bauer
IST Austria, TU München

Roland Kwitt
University of Salzburg, Austria

Abstract

1. Introduction

2. Related work

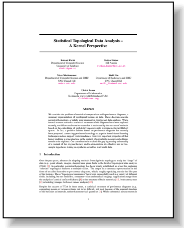
J. Reininghaus et al., ‘A stable multi-scale kernel for topological machine learning’

2013 2014 2015 2016 2017 2018 2019

Persistent homology

The second decade (almost)

A universal persistence scale space kernel


Statistical Topological Data Analysis – A Kernel Perspective

Roland Kwitt, Department of Computer Science, University of Salzburg
Stefan Huber, IST Austria
Marc Niethammer, Department of Computer Science and BRIC, UNC Chapel Hill
Weili Lin, Department of Radiology and BRIC, UNC Chapel Hill
Ulrich Bauer, Department of Mathematics, Technische Universität München (TUM)

Abstract

R. Kwitt et al., 'Statistical topological data analysis - A kernel perspective'

2013 2014 2015 2016 2017 2018 2019

Persistent homology

The second decade (almost)

Multidimensional fingerprints for molecules

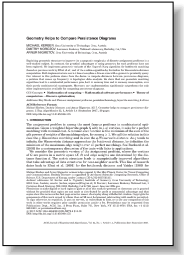
K. Xia and G. Wei, 'Multidimensional persistence in biomolecular data'

2013 2014 2015 2016 2017 2018 2019

Persistent homology

The second decade (almost)

Approximate metrics for persistence diagrams



Geometry Helps to Compare Persistence Diagrams

MICHAEL KERBER, Graz University of Technology, Graz, Austria
DMITRIY MOROZOV, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
ARNUR NIGMETOV, Graz University of Technology, Graz, Austria

Exploiting geometric structure to improve the asymptotic complexity of discrete assignment problems is a well-studied subject. In contrast, the practical advantages of using geometry for such problems have not been explored. We implement geometric variants of the Hopcroft-Karp algorithm for bottleneck matching (based on previous work by Efrat et al.) and of the auction algorithm by Bertsekas for Wasserstein distance computation. Both implementations use k-d trees to replace a linear scan with a geometric proximity query. Our interest in this problem stems from the desire to compute distances between persistence diagrams, a problem that comes up frequently in topological data analysis. We show that our geometric matching algorithms lead to a substantial performance gain, both in running time and in memory consumption, over their purely combinatorial counterparts. Moreover, our implementation significantly outperforms the only other implementation available for comparing persistence diagrams.

CCS Concepts: • **Mathematics of computing → Mathematical software performance;** • **Theory of computation → Discrete optimization;**

Additional Key Words and Phrases: Assignment problems, persistent homology, bipartite matching, k-d tree

ACM Reference Format:
Michael Kerber, Dmitriy Morozov, and Arnur Nigmetov. 2017. Geometry helps to compare persistence diagrams. J. Exp. Algorithmics 22, 1, Article 1.4 (September 2017), 20 pages.
DOI: http://dx.doi.org/10.1145/3064175

1. INTRODUCTION

The *assignment problem* is among the most famous problems in combinatorial optimization. Given a weighted bipartite graph G with $(n+n)$ vertices, it asks for a perfect matching with minimal cost. A common cost function is the minimum of the sum of the qth powers of weights of the matching edges, for some $q \geq 1$. We call the solution in this case the *q-Wasserstein matching* and its cost the *q-Wasserstein distance*. As q tends to infinity, the Wasserstein distance approaches the *bottleneck distance*, by definition the minimum of the maximum edge weight over all perfect matchings. See Burkard et al. [2009] for a contemporary discussion of the topic with links to applications.

We consider the geometric version of the assignment problem, where the vertices of G are points in a metric space (X, d) and edge weights are determined by the distance function d. The metric structure leads to asymptotically improved algorithms that take advantage of data structures for near-neighbor search. This line of research dates back to Efrat et al. [2001] for the bottleneck distance and Vaidya [1989] for

Michael Kerber and Arnur Nigmetov acknowledge support by the Max Planck Center for Visual Computing and Communication. Dmitriy Morozov is supported by Advanced Scientific Computing Research, Office of Science, U.S. Department of Energy, under Contract DE-AC02-05CH11231.
Authors' addresses: M. Kerber and A. Nigmetov, Institute of Geometry, Graz University of Technology, 8010 Graz, Austria; emails: {kerber, nigmetov}@tugraz.at; D. Morozov, Lawrence Berkeley National Lab, 1 Cyclotron Road, Mailstop 50R-3103, Berkeley, CA 94720; email: dmorozov@lbl.gov.

DOI: http://dx.doi.org/10.1145/3064175

M. Kerber et al., 'Geometry helps to compare persistence diagrams'

2013 2014 2015 2016 2017 2018 2019

Persistent homology

The second decade (almost)

Persistence-based features for histology classification

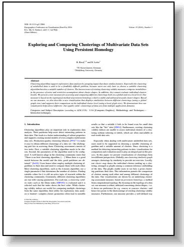
T. Qaiser et al., ‘Persistent homology for fast tumor segmentation in whole slide histology images’

2013 2014 2015 2016 2017 2018 2019

Persistent homology

The second decade (almost)

Persistence-based metric for cluster analysis



Exploring and Comparing Clusterings of Multivariate Data Sets Using Persistent Homology

B. Rieck[1,2] and H. Leitte[1]

[1]TU Kaiserslautern, Germany
[2]Heidelberg University, Germany

B. Rieck and H. Leitte, 'Exploring and comparing clusterings of multivariate data sets using persistent homology'

2013 2014 2015 2016 2017 2018 2019

Persistent homology

The second decade (almost)

Grid-based persistence diagram vectorisation


Journal of Machine Learning Research 18 (2017) 1-35 Submitted 7/16; Published 2/17

Persistence Images: A Stable Vector Representation of Persistent Homology

Henry Adams ADAMS@MATH.COLOSTATE.EDU
Tegan Emerson EMERSON@MATH.COLOSTATE.EDU
Michael Kirby KIRBY@MATH.COLOSTATE.EDU
Rachel Neville NEVILLE@MATH.COLOSTATE.EDU
Chris Peterson PETERSON@MATH.COLOSTATE.EDU
Patrick Shipman SHIPMAN@MATH.COLOSTATE.EDU
Department of Mathematics
Colorado State University
1874 Campus Delivery
Fort Collins, CO 80523-1874

Sofya Chepushtanova SOFYA.CHEPUSHTANOVA@WILKES.EDU
Department of Mathematics and Computer Science
Wilkes University
84 West South Street
Wilkes-Barre, PA 18766, USA

Eric Hanson ERIC.HANSON@TCU.EDU
Department of Mathematics
Texas Christian University
Box 298900
Fort Worth, TX 76129

Francis Motta MOTTA@MATH.DUKE.EDU
Department of Mathematics
Duke University
Durham, NC 27708, USA

Lori Ziegelmeier LZIEGEL1@MACALESTER.EDU
Department of Mathematics, Statistics, and Computer Science
Macalester College
1600 Grand Avenue
Saint Paul, MN 55105, USA

Editor: Michael Mahoney

Abstract

Many data sets can be viewed as a noisy sampling of an underlying space, and tools from topological data analysis can characterize this structure for the purpose of knowledge discovery. One such tool is persistent homology, which provides a multiscale description of the homological features within a data set. A useful representation of this homological information is a *persistence diagram* (PD). Efforts have been made to map PDs into spaces with additional structure valuable to machine learning tasks. We convert a PD to a finite-dimensional vector representation which we call a *persistence image* (PI), and prove the stability of this transformation with respect to small perturbations in the inputs. The

©2017 Adams, et al.

License: CC-BY 4.0, see https://creativecommons.org/licenses/by/4.0/. Attribution requirements are provided at http://jmlr.org/papers/v18/16-337.html.

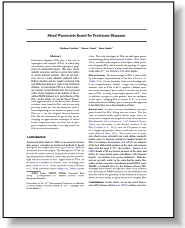
H. Adams et al., ‘Persistence images: A stable vector representation of persistent homology’

2013 2014 2015 2016 2017 2018 2019

Persistent homology

The second decade (almost)

Approximating the Wasserstein distance



M. Carrière et al., ‘Sliced Wasserstein Kernel for persistence diagrams’

2013 2014 2015 2016 2017 2018 2019

Persistent homology

The second decade (almost)

A persistence diagram layer for deep learning

C. Hofer et al., 'Deep learning with topological signatures'

Persistent homology

The second decade (almost)

Link prediction

arXiv:1811.04049v1 [cs.SI] 9 Nov 2018

Understanding and Predicting Links in Graphs: A Persistent Homology Perspective

Sumit Bhatia, Bapi Chatterjee, Deepak Nathani and Manohar Kaul

Abstract

Persistent Homology is a powerful tool in Topological Data Analysis (TDA) to capture topological properties of data succinctly at different spatial resolutions. For graphical data, shape, and structure of the neighborhood of individual data items (nodes) is an essential means of characterizing their properties. In this paper, we propose the use of persistent homology methods to capture structural and topological properties of graphs and use it to address the problem of link prediction. We evaluate our approach on seven different real-world datasets and offer directions for future work.

1 Introduction

A graph data structure representing pairwise relations or interactions among individuals or entities recurs in diverse real-world applications such as social and professional networks, biological phenomena such as protein-protein interactions [Coulomb et al., 2005], word co-occurrences [Newman, 2006], and citation and collaboration networks [Bhatia et al., 2012]. In all these applications, understanding how the network evolves and being able to predict the formation of new, hitherto non-existent links is an important problem in the knowledge discovery pipeline and has crucial applications such as predicting target genes for cancer research [Nagarajan et al., 2015], social network analysis, and recommendation systems.

S. Bhatia et al., 'Understanding and predicting links in graphs: A persistent homology perspective'

2013 2014 2015 2016 2017 2018 2019

Persistent homology

The second decade (almost)

Directed network analysis

Journal of Applied and Computational Topology (2018) 2:115–175
https://doi.org/10.1007/s41468-018-0020-6

CrossMark

A functorial Dowker theorem and persistent homology of asymmetric networks

Samir Chowdhury[1] · Facundo Mémoli[1,2]

Received: 4 January 2018 / Accepted: 4 September 2018 / Published online: 12 September 2018

Abstract
We study two methods for computing network features with topological underpinnings: the Rips and Dowker persistent homology diagrams. Our formulations work for general networks, which may be asymmetric and may have any real number as an edge weight. We study the sensitivity of Dowker persistence diagrams to asymmetry via numerous theoretical examples, including a family of highly asymmetric cycle networks that have interesting connections to the existing literature. In particular, we characterize the Dowker persistence diagrams arising from asymmetric cycle networks. We investigate the stability properties of both the Dowker and Rips persistence diagrams, and use these observations to run a classification task on a dataset comprising simulated hippocampal networks. Our theoretical and experimental results suggest that Dowker persistence diagrams are particularly suitable for studying asymmetric networks. As a stepping stone for our constructions, we prove a functorial generalization of a theorem of Dowker, after whom our constructions are named.

Keywords Directed networks · Functorial Dowker theorem · Cycle networks · Functorial Nerve Theorem

Mathematics Subject Classification 55U99 · 68U05 · 55N35

Contents

✉ Facundo Mémoli
memoli@math.osu.edu

Samir Chowdhury
chowdhury.57@osu.edu

1 Department of Mathematics, The Ohio State University, Columbus, OH, USA

2 Department of Computer Science and Engineering, The Ohio State University, Columbus, OH, USA

Springer

S. Chowdhury and F. Mémoli, 'A functorial Dowker theorem and persistent homology of asymmetric networks'

2013 2014 2015 2016 2017 2018 2019

Persistent homology

The second decade (almost)

Analysing the capacity of neural networks

On Characterizing the Capacity of Neural Networks using Algebraic Topology

William H. Guss Ruslan Salakhutdinov

Abstract

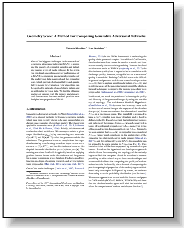
W. H. Guss and R. Salakhutdinov, ‘On characterizing the capacity of neural networks using algebraic topology’

2013 2014 2015 2016 2017 2018 2019

Persistent homology

The second decade (almost)

Comparing GAN feature spaces



Geometry Score: A Method For Comparing Generative Adversarial Networks

Valentin Khrulkov[1] Ivan Oseledets[1,2]

Abstract

1. Introduction

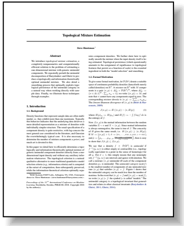
V. Khrulkov and I. Oseledets, ‘Geometry score: A method for comparing generative adversarial networks’

2013 2014 2015 2016 2017 2018 2019

Persistent homology

The second decade (almost)

Persistence for mixture estimation



Topological Mixture Estimation

Steve Huntsman

Abstract

1. Introduction

S. Huntsman, 'Topological mixture estimation'

2013 2014 2015 2016 2017 2018 2019

Persistent homology

The second decade (almost)

New kernel based on Fisher information metric


Persistence Fisher Kernel: A Riemannian Manifold Kernel for Persistence Diagrams

Tam Le
RIKEN Center for Advanced Intelligence Project, Japan

Makoto Yamada
Kyoto University, Japan
RIKEN Center for Advanced Intelligence Project, Japan

Abstract

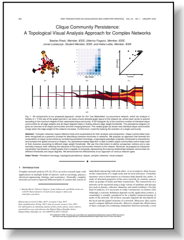
T. Le and M. Yamada, 'Persistence Fisher kernel: a Riemannian manifold kernel for persistence diagrams'

2013 2014 2015 2016 2017 2018 2019

Persistent homology

The second decade (almost)

Persistence for clique communities


IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 24, NO. 1, JANUARY 2018

Clique Community Persistence:
A Topological Visual Analysis Approach for Complex Networks

Bastian Rieck, *Member, IEEE*, Ulderico Fugacci, *Member, IEEE*,
Jonas Lukasczyk, *Student Member, IEEE*, and Heike Leitte, *Member, IEEE*

B. Rieck et al., ‘Clique Community Persistence: A topological visual analysis approach for complex networks’

2013 2014 2015 2016 2017 2018 2019

Persistent homology

The second decade (almost)

Deep sets for persistence diagrams


PersLay: A Simple and Versatile Neural Network Layer for Persistence Diagrams

Mathieu Carrière
Rabadan Lab
Columbia University
New York, US.

Frédéric Chazal
Datashape, Inria Saclay
Palaiseau, France.

Yuichi Ike
Fujitsu Laboratories, AI Lab
Tokyo, Japan.

Théo Lacombe
Datashape, Inria Saclay
Palaiseau, France.

Martin Royer
Datashape, Inria Saclay
Palaiseau, France.

Yuhei Umeda
Fujitsu Laboratories, AI Lab
Tokyo, Japan.

Abstract

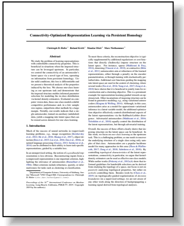
M. Carrière et al., 'PersLay: A Simple and Versatile Neural Network Layer for Persistence Diagrams'

2013 2014 2015 2016 2017 2018 2019

Persistent homology

The second decade (almost)

Topology-aware training with a new loss


Connectivity-Optimized Representation Learning via Persistent Homology

Christoph D. Hofer[1] Roland Kwitt[1] Mandar Dixit[2] Marc Niethammer[3]

Abstract

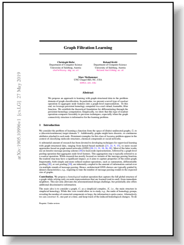
C. Hofer et al., 'Connectivity-optimized representation learning via persistent homology'

2013 2014 2015 2016 2017 2018 2019

Persistent homology

The second decade (almost)

Learning filtrations



arXiv:1905.10996v1 [cs.LG] 27 May 2019

Graph Filtration Learning

Christoph Hofer
Department of Computer Science
University of Salzburg, Austria
chofer@cosy.sbg.ac.at

Roland Kwitt
Department of Computer Science
University of Salzburg, Austria
Roland.Kwitt@sbg.ac.at

Marc Niethammer
UNC Chapel Hill, NC, USA
mn@cs.unc.edu

Abstract

We propose an approach to learning with graph-structured data in the problem domain of graph classification. In particular, we present a novel type of readout operation to aggregate node features into a graph-level representation. To this end, we leverage persistent homology computed via a real-valued, learnable, filter function. We establish the theoretical foundation for differentiating through the persistent homology computation. Empirically, we show that this type of readout operation compares favorably to previous techniques, especially when the graph connectivity structure is informative for the learning problem.

1 Introduction

We consider the problem of learning a function from the space of (finite) undirected graphs, $\mathcal{G}$, to a (discrete/continuous) target domain $\mathcal{Y}$. Additionally, graphs might have discrete, or continuous attributes attached to each node. Prominent examples for this class of *learning* problem appear in the context of classifying molecule structures, chemical compounds or social networks.

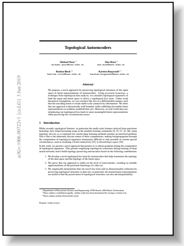
C. Hofer et al., 'Graph filtration learning'

2013 2014 2015 2016 2017 2018 2019

Persistent homology

The second decade (almost)

Optimising autoencoder representations


Topological Autoencoders

Michael Moor
Max Horn
Bastian Rieck
Karsten Borgwardt

Abstract

1 Introduction

M. Moor et al., ‘Topological autoencoders’

2013 2014 2015 2016 2017 2018 2019

Persistent homology

The second decade (almost)

Topology-based model selection

K. N. Ramamurthy et al., ‘Topological data analysis of decision boundaries with application to model selection’

2013 2014 2015 2016 2017 2018 2019

Persistent homology

The second decade (almost)

Neural network complexity analysis


Published as a conference paper at ICLR 2019

NEURAL PERSISTENCE: A COMPLEXITY MEASURE FOR DEEP NEURAL NETWORKS USING ALGEBRAIC TOPOLOGY

ABSTRACT

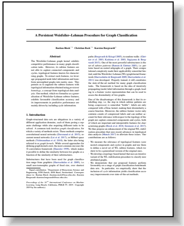
1 INTRODUCTION

B. Rieck et al., ‘Neural Persistence: A complexity measure for deep neural networks using algebraic topology’

Persistent homology

The second decade (almost)

Topological features for graph classification



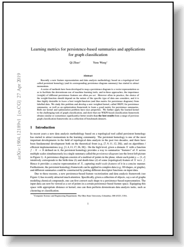
B. Rieck et al., 'A persistent Weisfeiler–Lehman procedure for graph classification'

2013 2014 2015 2016 2017 2018 2019

Persistent homology

The second decade (almost)

An optimised weight function for persistence images


arXiv:1904.12189v1 [cs.CG] 27 Apr 2019

Learning metrics for persistence-based summaries and applications for graph classification

Qi Zhao Yusu Wang

Abstract

1 Introduction

Q. Zhao and Y. Wang, ‘Learning metrics for persistence-based summaries and applications for graph classification’

2013 2014 2015 2016 2017 2018 2019

Three challenges for the future

We need to reduce the computational burden

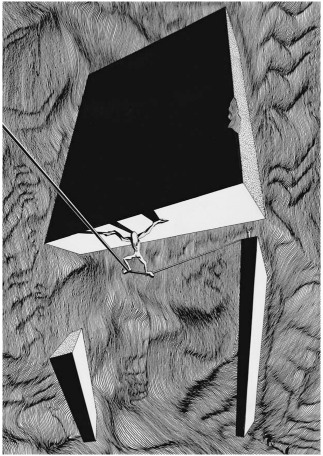


Image credit: Prof. A. T. Fomenko

One issue with the calculation is that, given the computational complexity of calculating $\mathfrak{R}_\epsilon(\cdot)$, we scale progressively worse with increasing batch size. In future work this could be mitigated by approximating the calculation of persistent homology or by exploiting recent advances in parallelising it.

M. Moor et al., 'Topological autoencoders'

We need to escape from Flatland

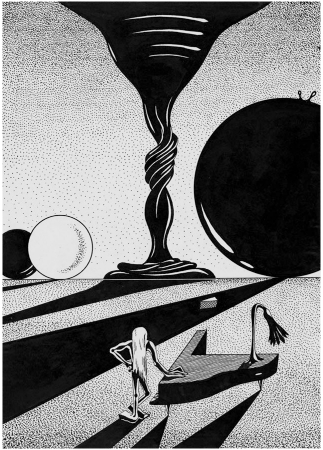


Image credit: Prof. A. T. Fomenko

While it would be theoretically possible to include higher-dimensional information about each layer G_k, [...], we focus on zero-dimensional information in this paper, because of the following advantages: i) the resulting values are easily interpretable [...], ii) previous research indicates that zero-dimensional topological information is already capturing a large amount of information, and iii) persistent homology calculations are highly efficient in this regime [...].

B. Rieck et al., 'Neural Persistence: A complexity measure for deep neural networks using algebraic topology'

We need proper architectures

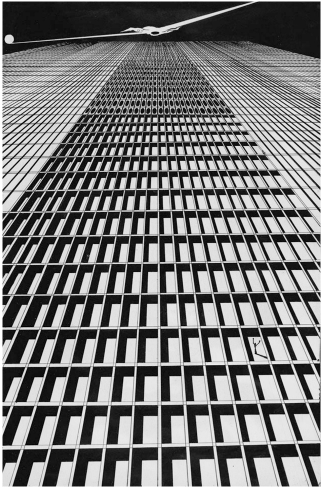


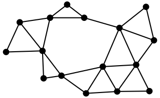
Image credit: Prof. A. T. Fomenko

So far, however, persistent homology is used in a passive manner, meaning that the function f mapping simplices to $\mathbb{R}$ is fixed and not informed by the learning task. Essentially, this degrades persistent homology to feature extraction step, where the obtained topological summaries are fed through a suitable vectorization scheme and passed to a classifier.

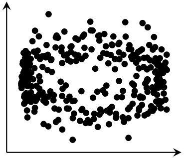
C. Hofer et al., 'Graph filtration learning'

Building intuition

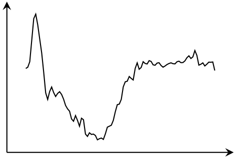
Pick suitable application domains



Graph classification



Feature space analysis



Time series classification

Graph classification

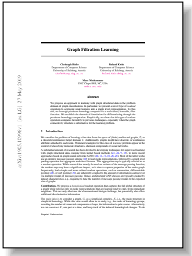
Some impulses

New architecture

Deep Learning with Topological Signatures

Christoph Hofer, Roland Kwitt, Marc Niethammer, Andreas Uhl

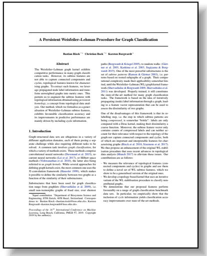
Filtration learning



Graph Filtration Learning

Christoph Hofer, Roland Kwitt, Marc Niethammer

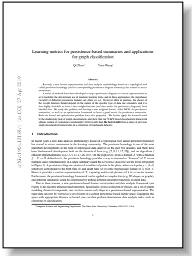
Subtree features



A Persistent Weisfeiler–Lehman Procedure for Graph Classification

Bastian Rieck, Christian Bock, Karsten Borgwardt

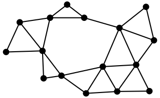
Weight functions



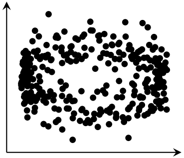
Learning metrics for persistence-based summaries and applications for graph classification

Qi Zhao, Yusu Wang

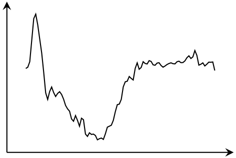
Pick suitable application domains



Graph classification

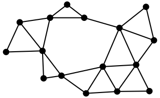


Feature space analysis

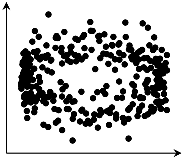


Time series classification

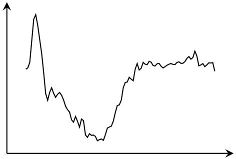
Pick suitable application domains



Graph classification



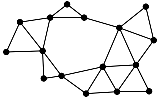
Feature space analysis



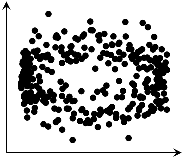
Time series classification

- V. Khrulkov and I. Oseledets, ‘Geometry score: A method for comparing generative adversarial networks’
- M. Moor et al., ‘Topological autoencoders’

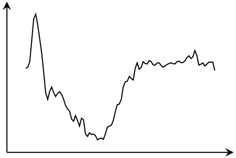
Pick suitable application domains



Graph classification



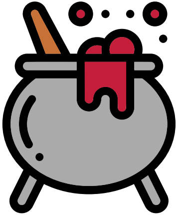
Feature space analysis



Time series classification

J. A. Perea et al., 'SW1PerS: Sliding windows and 1-persistence scoring; discovering periodicity in gene expression time series data'

What to avoid

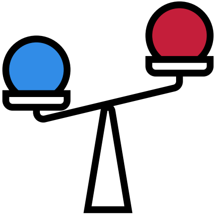


Round about the cauldron go;
In the persistent entrails throw.
Diagram that with many a pair
Makes the network look less bare.

Double, double toil and trouble;
Fire burn and cauldron bubble.

Persistent homology should *not* become another 'ingredient' in our networks that we do not understand.

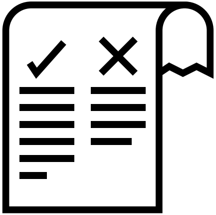
How to write successful papers in topological machine learning



Choose suitable comparison partners. Do *not* restrict yourself to TDA-based techniques but choose the *best* techniques you can find (including TDA baselines).

Good example: Q. Zhao and Y. Wang, ‘Learning metrics for persistence-based summaries and applications for graph classification’

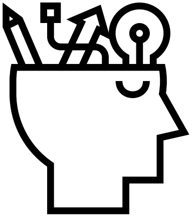
How to write successful papers in topological machine learning



Show the *benefits* of persistent homology or TDA. Why TDA and not something else?

Good example: V. Khrulkov and I. Oseledets, 'Geometry score: A method for comparing generative adversarial networks'

How to write successful papers in topological machine learning

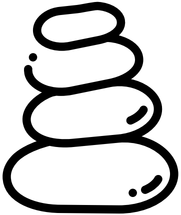


Explain TDA. What is the meaning of topological features for a particular data set?

Good example: B. Rieck et al., ‘A persistent Weisfeiler–Lehman procedure for graph classification’

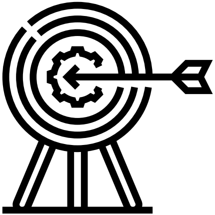
Our tasks

Standards



We need to *harmonise* and *standardise* our methods to encourage sharing and the creation of frameworks.

Benchmark data sets



We need benchmark data sets against which we can test all our methods; plus, benchmarks help us present compelling examples.

Conclusion

Many interesting challenges lie ahead! If you want to help, I would love to hear from you:

bastian.rieck.me
bastian.rieck@bsse.ethz.ch
, Pseudomanifold

Thank you very much for your attention!

Acknowledgements

Research

Christian Bock, Karsten Borgwardt, Thomas Gumbsch, Max Horn, Michael Moor, Matteo Togninalli

Illustrations

The illustrations for the ‘Challenges’ have been created by Prof. A. T. Fomenko for his book *Homotopical Topology*. He kindly gave permission to use them in this talk.

Icons

The icons were originally created by Freepik and Eucalyp from Flaticon. They have been slightly modified in some cases.