AFRL-AFOSR-VA-TR-2016-0333

Statistical inferences from the topology of complex networks

John Holcomb
CLEVELAND STATE UNIV OH
2121 EUCLID AVE
CLEVELAND, OH 44115 - 2226

10/11/2016
Final Report

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/RTA2

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
04-10-2016	Final Report	3/01/13 - 08/31/16

4. TITLE AND SUBTITLE
Statistical Inferences from the Topology of Complex Networks

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-13-1-0115

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Bubenik, Peter - University of Florida
Holcomb, John - Cleveland State University

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Cleveland State University - 2121 Euclid Ave., Cleveland, OH 44115-2226
University of Florida - 219 Grinter Hall, Gainsville, FL 32611-5500

8. PERFORMING ORGANIZATION REPORT NUMBER
200000171

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
USAF, AFRL
AF Office of Scientific Research
875 North Randolph Street, RM 3112
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
AFOSR/PKR2

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
FA9550

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Topological data analysis provides machinery for summarizing the topology of complex data. The main goal of this project was to develop a new summary compatible with statistics and machine learning. This goal was met with the development of a new summary, the "persistence landscape." This summary is stable, does not lose any information, has continuous and discrete versions, and obeys a strong law of large numbers and a central limit theorem. All of the standard tools in statistics and machine learning are available for subsequent analysis. For example, one can easily calculate averages and differences, apply principal component analysis and support vector machines, or feed these results into a neural network. The secondary goal of the project was to help place Topological Data Analysis on a firmer mathematical foundation, strengthening its connections to mathematics and making it easier for researchers to leverage mathematical results for analyzing complex data and complex networks. This goal has been met with the development a very general framework for topological data analysis. Both of these developments have led to much work by other researchers.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(Include area code)*

Reset

INSTRUCTIONS FOR COMPLETING SF 298

1. REPORT DATE. Full publication date, including day, month, if available. Must cite at least the year and be Year 2000 compliant, e.g. 30-06-1998; xx-06-1998; xx-xx-1998.

2. REPORT TYPE. State the type of report, such as final, technical, interim, memorandum, master's thesis, progress, quarterly, research, special, group study, etc.

3. DATES COVERED. Indicate the time during which the work was performed and the report was written, e.g., Jun 1997 - Jun 1998; 1-10 Jun 1996; May - Nov 1998; Nov 1998.

4. TITLE. Enter title and subtitle with volume number and part number, if applicable. On classified documents, enter the title classification in parentheses.

5a. CONTRACT NUMBER. Enter all contract numbers as they appear in the report, e.g. F33615-86-C-5169.

5b. GRANT NUMBER. Enter all grant numbers as they appear in the report, e.g. AFOSR-82-1234.

5c. PROGRAM ELEMENT NUMBER. Enter all program element numbers as they appear in the report, e.g. 61101A.

5d. PROJECT NUMBER. Enter all project numbers as they appear in the report, e.g. 1F665702D1257; ILIR.

5e. TASK NUMBER. Enter all task numbers as they appear in the report, e.g. 05; RF0330201; T4112.

5f. WORK UNIT NUMBER. Enter all work unit numbers as they appear in the report, e.g. 001; AFAPL30480105.

6. AUTHOR(S). Enter name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. The form of entry is the last name, first name, middle initial, and additional qualifiers separated by commas, e.g. Smith, Richard, J, Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES). Self-explanatory.

8. PERFORMING ORGANIZATION REPORT NUMBER. Enter all unique alphanumeric report numbers assigned by the performing organization, e.g. BRL-1234; AFWL-TR-85-4017-Vol-21-PT-2.

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES). Enter the name and address of the organization(s) financially responsible for and monitoring the work.

10. SPONSOR/MONITOR'S ACRONYM(S). Enter, if available, e.g. BRL, ARDEC, NADC.

11. SPONSOR/MONITOR'S REPORT NUMBER(S). Enter report number as assigned by the sponsoring/ monitoring agency, if available, e.g. BRL-TR-829; -215.

12. DISTRIBUTION/AVAILABILITY STATEMENT. Use agency-mandated availability statements to indicate the public availability or distribution limitations of the report. If additional limitations/ restrictions or special markings are indicated, follow agency authorization procedures, e.g. RD/FRD, PROPIN, ITAR, etc. Include copyright information.

13. SUPPLEMENTARY NOTES. Enter information not included elsewhere such as: prepared in cooperation with; translation of; report supersedes; old edition number, etc.

14. ABSTRACT. A brief (approximately 200 words) factual summary of the most significant information.

15. SUBJECT TERMS. Key words or phrases identifying major concepts in the report.

16. SECURITY CLASSIFICATION. Enter security classification in accordance with security classification regulations, e.g. U, C, S, etc. If this form contains classified information, stamp classification level on the top and bottom of this page.

17. LIMITATION OF ABSTRACT. This block must be completed to assign a distribution limitation to the abstract. Enter UU (Unclassified Unlimited) or SAR (Same as Report). An entry in this block is necessary if the abstract is to be limited.

Final Report
Statistical Inferences from the Topology of Complex Networks
FA9550-13-1-0115

Awarded to PI: Peter Bubenik
Current PI: John Holcomb

October 4, 2016

The project "Statistical Inferences from the Topology of Complex Networks" was funded by the AFOSR under the Complex Networks program, for the period March 1, 2013 to February 29, 2016. It was granted a No Cost Extension from March 1, 2016 to August 31, 2016. Here we offer a final report on the results of this project for the duration of the grant, from March 1, 2013 to August 31, 2016.

Change in PI

The project was awarded to Dr. Peter Bubenik at Cleveland State University. In August 2015, Dr. Bubenik moved to the University of Florida. Consequently, the PI on the project was changed to Dr. John Holcomb, the Chair of the Department of Mathematics at Cleveland State University. The remaining balance on the grant was sub-awarded to Dr. Bubenik at the University of Florida, where he continued to work on the project. All references to PI below are to Dr. Peter Bubenik.

Summary of the research

Main goals

As described in the abstract of the proposal for this project, its goals were "to develop and study a new topological descriptor that is designed for statistical

inference, and to use it for that purpose," and "to develop this topological machinery from a more abstract point of view, providing a better framework for studying its stability and for extending the scope of this technology."

Topological data analysis provides machinery for summarizing the topology of complex data as a "barcode" or "persistence diagram". While these have been successful tools for visualization, they are unsuitable for further statistical analysis or machine learning. The main goal of this project was to develop a new summary compatible with statistics and machine learning. This goal was met with the development of a new summary, the "persistence landscape". This summary is stable, does not lose any information, has continuous and discrete versions, and obeys a strong law of large numbers and a central limit theorem. The main results were published in the Journal of Machine Learning Research in a paper titled "Statistical topological data analysis using persistence landscapes" [4]. It is a functional summary which may be viewed as a point in a vector space (or more precisely, a point in a Hilbert space), and all of the standard tools in statistics and machine learning are available for subsequent analysis. For example, one can easily calculate averages and differences, and apply principal component analysis and support vector machines, or feed these results into a neural network.

The secondary goal of the project was to help place Topological Data Analysis on a firmer mathematical foundation, strengthening its connections to mathematics and making it easier for researchers to leverage mathematical results for analyzing complex data and complex networks. This goal was met with the publication of the paper (with J.A. Scott) "Categorification of persistent homology" [7] in the journal Discrete and Computational Geometry and the paper "Metrics for Generalized Persistence Modules" (with J.A. Scott and V. de Silva) in the journal Foundations of Computational Mathematics [5]. These papers develop a very general framework for topological data analysis.

Extensions of original goals

With the main goals achieved a number of extensions of these goals were pursued.

The topological summary, the "persistence landscape" developed in this project was validated by demonstrating that it could be combined with statistical inference and machine learning in a biological application. The resulting paper, "Using persistent homology and dynamical distances to analyze pro-

tein binding" (with V. Kovacev-Nikolic, D. Nikolic, and G. Heo) appeared in Statistical Applications in Genetics and Molecular Biology [18].

Together with P. Dlotko, the PI developed efficient algorithms for constructing the persistence landscape and for combining it with statistical analysis. The resulting article, "A persistence landscapes toolbox for topological statistics" is in the Journal of Symbolic Computation [6].

Together with P. Bendich, the PI has developed a framework and algorithm for stabilizing the location of topological features and also to stabilize topological computations with respect to choices of parameters. The paper "Stabilizing the output of persistent homology computations" has been submitted [3].

Together with V. de Silva and V. Nanda, the PI has studied the geometry of the algebraic objects of study in Topological Data Analysis. The resulting paper, "Higher interpolation and extension of persistence modules" is undergoing peer review.

Impact on the community

The main contribution of this project, the persistence landscape, is perhaps the most influential development in this research area in the past few years. The paper [4] has inspired considerable theoretical research and is starting to be used in a wide variety of applications. The bootstrap has been applied to provide confidence bands for the persistence landscape [11, 10]. The persistence landscape has also inspired a number of other linear topological summaries [11, 22, 8, 23, 1, 19, 12, 13, 25, 2]. The persistence landscape has been used to study brain images [24], fluid dynamics [16], brain EEG data [26], complex networks [9], and phase transitions [17]. It has also been recently combined with Neural Networks to study audio signals [20]. According to Google Scholar this paper has already been cited 70 times.

The secondary goal of the project was the development of a general framework for topological data analysis, which was given in [7] and [5]. This framework has already been used by other researchers to develop algorithms and prove properties in concrete settings such as those for Reeb graphs [14, 21]. According the Google Scholar these papers have 42 and 13 citations respectively.

Software development

The algorithms for persistence landscapes and statistical inference presented in [6] have been implemented as "The Persistence Landscape Toolbox" and the code is publicly available [15].

Dissemination of results

During this project the PI gave 26 invited lectures describing its results. The venues for these talks included research institutes and leading universities in the United States, Canada, Mexico, the United Kingdom, Germany, Denmark, Poland and Japan.

Broader impacts

The PI has made efforts to increase the broader impact of the project. He founded and serves as Director of the Applied Algebraic Topology Research Network, a network with close to 300 members that is funded by the NSF through the Institute of Mathematics and its Applications (IMA). He also accepted a position as Associate Editor at the new (Society for Industrial and Applied Mathematics) SIAM Journal on Applied Algebra and Geometry. In addition he was a member of the Scientific Committee for the main conference in this subject, "Applied Topology: Computation, Methods, and Science," held in Turin, Italy, in July 2016.

The PI has given outreach talks at the NASA Glenn Research Center's Summer Intern Seminar, and to the University of Florida Graduate Mathematics Association Colloquium. He was also the featured speaker at a Summer School organized by the Mathematical Association of America entitled "Big Data on the Great Plains."

At Cleveland State University and the University of Florida (UF) he has been using the results of this project to teach undergraduate and graduate students. At UF, he has incorporated TDA into the graduate topology course and he started a Student Applied Topology seminar.

At UF, the PI has also been training Highly Qualified Personnel. These will be future researchers and data analysts and this training will aid the competitiveness of the United States. He is advising one postdoctoral researcher, Dr. Michael Catanzaro, two Ph.D. candidates, Alexander Wagner

and Leo Betthauser, and two undergraduate students, Benjamin Whittle and Dhruv Patel.

References

[1] Aaron Adcock, Erik Carlsson, and Gunnar Carlsson. The ring of algebraic functions on persistence bar codes. *Homology Homotopy Appl.*, 18(1):381–402, 2016.

[2] Rushil Anirudh, Vinay Venkataraman, Karthikeyan Natesan Ramamurthy, and Pavan Turaga. A riemannian framework for statistical analysis of topological persistence diagrams. 05 2016, 1605.08912.

[3] Paul Bendich and Peter Bubenik. Stabilizing the output of persistent homology computations. 12 2015, 1512.01700.

[4] Peter Bubenik. Statistical topological data analysis using persistence landscapes. *Journal of Machine Learning Research*, 16:77–102, 2015.

[5] Peter Bubenik, Vin de Silva, and Jonathan Scott. Metrics for Generalized Persistence Modules. *Found. Comput. Math.*, 15(6):1501–1531, 2015.

[6] Peter Bubenik and Pawel Dlotko. A persistence landscapes toolbox for topological statistics. *Journal of Symbolic Computation*, 78:91 – 114, 2017. Algorithms and Software for Computational Topology.

[7] Peter Bubenik and Jonathan A. Scott. Categorification of persistent homology. *Discrete Comput. Geom.*, 51(3):600–627, 2014.

[8] Mathieu Carrière, Steve Y. Oudot, and Maks Ovsjanikov. Stable topological signatures for points on 3d shapes. *Computer Graphics Forum*, 34(5):1–12, 2015.

[9] Corrie Jacobien Carstens. *Topology of Complex Networks: Models and Analysis*. PhD thesis, RMIT University, 2016.

[10] Frédéric Chazal, Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, Aarti Singh, and Larry Wasserman. On the bootstrap for persistence diagrams and landscapes. *Modeling and Analysis of Information Systems*, 20(6):96–105, 2014, 1311.0376.

[11] Frédéric Chazal, Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, and Larry Wasserman. Stochastic convergence of persistence landscapes and silhouettes. *J. Comput. Geom.*, 6(2):140–161, 2015.

[12] Yen-Chi Chen, Daren Wang, Alessandro Rinaldo, and Larry Wasserman. Statistical analysis of persistence intensity functions. 10 2015, 1510.02502.

[13] Sofya Chepushtanova, Tegan Emerson, Eric Hanson, Michael Kirby, Francis Motta, Rachel Neville, Chris Peterson, Patrick Shipman, and Lori Ziegelmeier. Persistence images: An alternative persistent homology representation. 07 2015, 1507.06217.

[14] Vin de Silva, Elizabeth Munch, and Amit Patel. Categorified Reeb graphs. *Discrete Comput. Geom.*, 55(4):854–906, 2016.

[15] Pawel Dlotko. The persistence landscape toolbox. `https://www.math.upenn.edu/~dlotko/persistenceLandscape.html`.

[16] Paweł Dłotko and Thomas Wanner. Topological microstructure analysis using persistence landscapes. *Physica D: Nonlinear Phenomena*, 334:60 – 81, 2016. Topology in Dynamics, Differential Equations, and Data.

[17] Irene Donato, Matteo Gori, Marco Pettini, Giovanni Petri, Sarah De Nigris, Roberto Franzosi, and Francesco Vaccarino. Persistent homology analysis of phase transitions. *Phys. Rev. E*, 93:052138, May 2016.

[18] Violeta Kovacev-Nikolic, Peter Bubenik, Dragan Nikolić, and Giseon Heo. Using persistent homology and dynamical distances to analyze protein binding. *Stat. Appl. Genet. Mol. Biol.*, 15(1):19–38, 2016.

[19] Genki Kusano, Kenji Fukumizu, and Yasuaki Hiraoka. Persistence weighted gaussian kernel for topological data analysis. In *Proceedings of the 33rd International Conference on Machine Learning*, volume 48. JMLR: W&CP, 2016.

[20] Jen-Yu Liu, Shyh-Kang Jeng, and Yi-Hsuan Yang. Applying topological persistence in convolutional neural network for music audio signals. 08 2016, 1608.07373.

[21] Elizabeth Munch and Bei Wang. Convergence between categorical representations of reeb space and mapper. 12 2015, 1512.04108.

[22] Jan Reininghaus, Stefan Huber, Ulrich Bauer, and Roland Kwitt. A stable multi-scale kernel for topological machine learning. In *Proc. 2015 IEEE Conf. Comp. Vision & Pat. Rec. (CVPR '15)*, 2015.

[23] Vanessa Robins and Katharine Turner. Principal component analysis of persistent homology rank functions with case studies of spatial point patterns, sphere packing and colloids. *Physica D: Nonlinear Phenomena*, pages –, 2016.

[24] Bernadette J. Stolz, Heather A. Harrington, and Mason A. Porter. Persistent homology of time-dependent functional networks constructed from coupled time series. 05 2016, 1605.00562.

[25] Sara Kalisnik Verovsek. Tropical coordinates on the space of persistence barcodes. 04 2016, 1604.00113.

[26] Y. Wang, H. Ombao, and M. K. Chung. Topological seizure origin detection in electroencephalographic signals. In *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pages 351–354, April 2015.

AFOSR Deliverables Submission Survey

Response ID:7060 Data

1.

Report Type

Final Report

Primary Contact Email

Contact email if there is a problem with the report.

peter.bubenik@gmail.com

Primary Contact Phone Number

Contact phone number if there is a problem with the report

3522942342

Organization / Institution name

University of Florida

Grant/Contract Title

The full title of the funded effort.

Statistical Inferences from the Topology of Complex Networks

Grant/Contract Number

AFOSR assigned control number. It must begin with "FA9550" or "F49620" or "FA2386".

FA9550-13-1-0115

Principal Investigator Name

The full name of the principal investigator on the grant or contract.

Peter Bubenik

Program Officer

The AFOSR Program Officer currently assigned to the award

James Lawton

Reporting Period Start Date

03/01/2013

Reporting Period End Date

08/31/2016

Abstract

Topological data analysis provides machinery for summarizing the topology of complex data. The main goal of this project was to develop a new summary compatible with statistics and machine learning. This goal was met with the development of a new summary, the "persistence landscape." This summary is stable, does not lose any information, has continuous and discrete versions, and obeys a strong law of large numbers and a central limit theorem. All of the standard tools in statistics and machine learning are available for subsequent analysis. For example, one can easily calculate averages and differences, apply principal component analysis and support vector machines, or feed these results into a neural network. The secondary goal of the project was to help place Topological Data Analysis on a firmer mathematical foundation, strengthening its connections to mathematics and making it easier for researchers to leverage mathematical results for analyzing complex data and complex networks. This goal has been met with the development a very general framework for topological data analysis. Both of these developments have led to much work by other researchers.

Distribution Statement

This is block 12 on the SF298 form.

Distribution A - Approved for Public Release

Explanation for Distribution Statement

If this is not approved for public release, please provide a short explanation. E.g., contains proprietary information.

SF298 Form

Please attach your SF298 form. A blank SF298 can be found here. Please do not password protect or secure the PDF The maximum file size for an SF298 is 50MB.

afosr-sf298 bubenik.pdf

Upload the Report Document. File must be a PDF. Please do not password protect or secure the PDF . The maximum file size for the Report Document is 50MB.

final_report_2016.pdf

Upload a Report Document, if any. The maximum file size for the Report Document is 50MB.

Archival Publications (published) during reporting period:

Peter Bubenik and Jonathan A. Scott. Categorification of persistent homology, Discrete and Computational Geometry 51 (2014), no. 3, 600–627.
Peter Bubenik. Statistical topological data analysis using persistence landscapes, Journal of Machine Learning Research 16 (2015), 77–102.
Peter Bubenik, Vin de Silva, and Jonathan Scott. Metrics for generalized persistence modules, Foundations of Computational Mathematics 15 (2015), no.6, 1501–1531.
Violeta Kovacev-Nikolic, Peter Bubenik, Dragan Nikolic, and Giseon Heo. Using persistent homology and dynamical distances to analyze protein binding, Statistical Applications in Genetics and Molecular Biology 15 (2016) no. 1, 19–38.
Peter Bubenik and Pawel Dlotko. A persistence landscapes toolbox for topological statistics, Journal of Symbolic Computation 78 (2017), 91–114.

New discoveries, inventions, or patent disclosures:
Do you have any discoveries, inventions, or patent disclosures to report for this period?

No

Please describe and include any notable dates

Do you plan to pursue a claim for personal or organizational intellectual property?

Changes in research objectives (if any):

Change in AFOSR Program Officer, if any:

Program Officer at time of award: Dr. Robert Bonneau
Current Program Officer: Dr. James H Lawton

Extensions granted or milestones slipped, if any:

No Cost Extension, 1 Mar 2016 - 31 Aug 2016.

AFOSR LRIR Number

LRIR Title

Reporting Period

Laboratory Task Manager

Program Officer

Research Objectives

Technical Summary

Funding Summary by Cost Category (by FY, $K)

	Starting FY	FY+1	FY+2
Salary			
Equipment/Facilities			
Supplies			
Total			

Report Document

Report Document - Text Analysis

Report Document - Text Analysis

Appendix Documents

2. Thank You

E-mail user

Oct 04, 2016 14:39:03 Success: Email Sent to: peter.bubenik@gmail.com