Julien Tierny, Ph.D.,
CNRS Research Director
Sorbonne Université
Computer Science Department (LIP6)
4 Place Jussieu
75 005 Paris, France
Phone: (33) 1 44 27 88 25
Email: julien.tierny@sorbonne-universite.fr
Web: https://julien-tierny.github.io/

Evaluation report for Davide Gurnari's thesis

Dear evaluation committee,

It is my pleasure to present a review of Mr. Davide Gurnari's Ph.D. thesis, entitled "New Shape Descriptors for Topological Data Analysis".

Topic

Davide Gurnari's Ph.D. work deals with the development of new topological descriptors for the *(i)* quantitative and *(ii)* qualitative analysis of high-dimensional point cloud data. Specifically, for each of these two categories of analyses, Davide Gurnari introduced two contributions. On one hand, he presented a generalization of Euler characteristic curves to multiparameter filtrations, as well as a new computational approach for extracting relevant representative cycles of persistent homology classes. On the other hand, he introduced practical variants of the Mapper construction as well as a new representation, the Cluster Graph, enabling the visual analysis of the relationships between clusters within the input data. Each of these contributions has been validated with real-life data, with software implementations made publicly available.

Manuscript

The manuscript itself is 108 pages long (without references). It is entirely written in English. It includes a short introduction followed by five chapters.

The introduction motivates the research presented in the thesis and provides a brief summary of the manuscript, which greatly helps its reading.

Chapter 2 presents a summary of the theoretical background required for the understanding of the rest of the manuscript. It provides a concise yet clear formalization of key concepts related to persistent homology.

Chapter 3 presents the first contribution of the work dealing with the extension of Euler characteristic curves (ECC) to multiparameter filtrations. First, this chapter introduces a stability result for these signatures in the case of univariate filtrations, specifically, relative to the Wasserstein distance between the corresponding persistence diagrams. This result further motivates the use of ECCs in practice. Second, a generalization of these signatures, called the Euler characteristic profiles (ECP), is introduced for multiparameter filtrations

and its practical stability is discussed. Third, an algorithm is presented for the computation of ECCs and EPPs. In particular, this algorithm processes each data point independently, which allows for a parallel computation. Experiments within a shared-memory parallelism context illustrates the practical gains of this parallelization strategy. Last, this chapter concludes on the practical application of these descriptors in classification tasks on real-life data from biology.

Chapter 4 presents the second contribution of the thesis, dealing with the extraction of relevant representative cycles for persistent homology classes. For this, the approach relies on the notion of harmonic homology for the unambiguous identification of a unique representative cycle per homology class. The uniqueness of such a cycle is of particular interest in practice as the examples of representative cycles traditionally generated during persistent homology computation can be challenging to interpret (due to their instability in geometry or even topology, e.g., number of connected components). This chapter presents the first algorithm for the computation of such unique cycles, called harmonic cycles. Moreover, some simplex-wise weights derived from the notion of "essentiality" of harmonic cycles are introduced and their relevance is demonstrated in an application to machine learning tasks on real-life biological data.

Chapter 5 presents the third contribution of the thesis, dealing with various extensions of the Mapper construction for the qualitative, visual analysis of high-dimensional data, including an extension of the Ball mapper (depicting the action of input isometries), an extension to high-dimensional lens functions for the Ball mapper as well as a Mapper-derived representation for studying correlations between datasets. The practical relevance of each of these extensions is discussed, in particular in an application to the analysis of polynomials in knot theory.

Chapter 6 presents the last contribution of the thesis, dealing with a new representation, called the ClusterGraph, characterizing the structural relations between clusters in high-dimensional datasets. This new representation is then evaluated in practice, in particular by comparing it to standard dimensionality reduction techniques, for the visual inspection of real-life data.

Appreciation

Mr Davide Gurnari's delivered a remarkably well written and easy-to-read manuscript, with a very clear and motivated structure, which helps appreciate his approach through this research thesis. The manuscript also contains multiple clear illustrations, which further helps its reading. The document describes valuable contributions to the field of Topological Data Analysis (both in its quantitative and qualitative aspects), which have been validated with publications in international journals of very good quality. In particular, I found the development of tools based on harmonic homology particularly refreshing and intriguing and I suspect that this line of work may inspire other researchers. In terms of structural coherence, it may have been beneficial to include a concluding chapter, with a reflective discussion on the work and its perspectives. Moreover, it would have been interesting to also include a formalization of the Mapper algorithm within the background section, in particular to highlight its roots in the Reeb graph. Despite these editorial remarks, I believe that Davide's manuscript was very pleasant to read and insightful.
Finally, I would like to stress the remarkable additional efforts made by Davide to make publicly available both the data he analyzed and the implementations he has written. Such an effort is unfortunately often under-appreciated in academia, although it is an absolute necessity in modern reproducible science. Thanks to his efforts, Davide's work can be easily reproduced by others, which in itself is also a valuable contribution for our research community.

For these reasons, I "deem the thesis as sufficient to grant a Ph.D." and I am happy to support its defense.

Julien Tierny
Paris,
January 10, 2025