Steve Oudot
Research Scientist, Inria
Part-Time Professor, École Polytechnique
Alan Turing Bldg., office 2105
École polytechnique
91120 Palaiseau, France
Phone: (+33) 174 854 216
email: `steve.oudot@inria.fr`

April 10, 2018

Subject: Recommendation letter for Mathieu Carrière.

Dear members of the selection committee,

It is with great pleasure that I recommend my former PhD student Mathieu Carrière for the Prix de la Chancellerie. Our research group (the DataShape team at Inria, see `http://team.inria.fr/datashape/`) has trained several excellent students in the past years, some of whom received prestigious awards (e.g. the Gilles Kahn prize) for their PhD work. To me, Mathieu stands out as the best one among them.

Mathieu's work lies within the field of applied topology and focuses primarily on *Topological Data Analysis* (TDA), a young research topic at the interface between pure mathematics, applied mathematics, and computer science. The general aim of TDA is to integrate concepts and tools coming from algebraic topology into methods for data analysis. One of its most concrete outcomes has been the introduction of new families of descriptors for data that are simple to compute and to compare, in arbitrary dimensions (actually in arbitrary metric or dissimilarity spaces), with provable invariance and stability properties. Such descriptors can serve e.g. as a visualization tool in data exploration, or as features for machine learning. They have been used successfully in a variety of applications coming from astrophysics, material sciences, structural biology, dynamical systems, or healthcare, to name a few. See e.g. S. Oudot, *Persistence Theory: From Quiver Representations to Data Analysis* (AMS Surveys and Monographs, number 209, 2015) for a detailed list of applications and an introduction to the subject.

Today, TDA is arousing a lot of interest among the data Science community at large. However, at the time when Mathieu's PhD started, the subject had developed mostly independently from the rest of the literature. Indeed, several major bottlenecks needed to be lifted in order to effectively integrate TDA tools into standard data analysis pipelines: (a) the interpretability of the descriptors; (b) the scalability of their computations and comparisons; (c) the lack of variety in the topological constructions that lead to these descriptors; (d) the lack of a Hilbert (or even linear) structure on the space in which they live. Mathieu's contributions have shed new light on some these questions, and they span all aspects of TDA. Let me highlight a few of them:

(1) Addressing bottleneck (c) head on, Mathieu proposed novel topological constructions that

yield new families of descriptors with enough flexibility to characterize either an entire dataset or a subpopulation therein. This makes it possible e.g. to compare observations within a same dataset or across datasets. We leveraged this flexibility in an application coming from computer graphics, where the goal is to segment 3d shapes automatically from examples. We showed that adding our topological descriptors on top of the classical set of geometric descriptors greatly improves the performances of the state-of-the-art segmentation methods, using fewer examples. On the mathematical side, the invariance and stability properties of our descriptors were proven formally using a combination of tools coming from algebraic topology on the one hand, metric geometry on the other hand.

(2) Addressing bottlenecks (b) and (d), Mathieu proposed effective vectorization methods and kernels for the descriptors of TDA, which allow for fast comparisons on the one hand, for their use as features in machine learning on the other hand. The main challenge faced in this work was to guarantee that the corresponding feature maps are metric-preserving, so that the stability and discrimination properties of the descriptors still hold after the embedding into Hilbert space. Mathieu's approach was to treat the descriptors as discrete signed measures, a novel viewpoint on the problem that allowed him to leverage vectorizations and kernels from the literature in optimal transport, and to adapt them to our needs. The adaptation per se was non-trivial because the metric between our descriptors is similar to an unbalanced Wasserstein distance. The originality of this work truly lies in the connection between TDA and optimal transport, which was seminal and has since inspired new developments, including by ourselves.

(3) Addressing bottleneck a), Mathieu connected the descriptors of TDA to certain families of graphs used for scientific visualization, in particular *Reeb graphs* and their discrete approximations (e.g. *Mapper graphs*). The benefit of this connection is twofold: on the one hand, it allows users to interpret our descriptors as encodings of the structure of such graphs, which they can visualize and understand visually; on the other hand, it provides a sound mathematical framework to analyze the stability properties of such graphs with respect to perturbations of the originating real-valued functions or of their domains. Following up on the same line of work, Mathieu then set up a statistical framework for analyzing the bias and variance in Reeb graph estimators. He used this framework to derive deviation bounds and confidence regions for some of the Reeb graph estimators. For this Mathieu studied the measurability of the mapping from a graph to its corresponding descriptor, then he used push-forwards into the space of descriptors to carry out the statistical analysis. Down the road, this work allowed us to propose automatic parameter selection methods for Reeb graph estimators, a long-awaited feature for applications. We leveraged this approach in a recent collaboration with researchers at EPFL on the detection and classification of cancer subtypes from genome expressions, a contribution soon to be submitted to *Nature Methods*.

Put together, these contributions constitute a key step towards the integration of TDA into the core of the data Science literature. They impact all aspects of TDA, from mathematical foundations to algorithmic developments, to statistical analysis, to combination with machine learning, to applications. This is a perfect illustration of TDA as a transversal topic.

Throughout his PhD, Mathieu showed a strong leadership and a lot of independent thinking, proposing most of the key ideas that have led to the aforementioned advances, and following up very quickly on the other ideas. He had the leading role in all our projects together, and he also connected to other researchers (e.g. Prof. U. Bauer from TUM) for follow-up work. His skills as an applied mathematician and as a computer scientist are undeniably of the strongest type, with an innovative and rigorous mind, two essential qualities for a researcher. He is able to translate high-level ideas into formal statements and proofs very naturally. He is also very well organized, and I am really impressed by how well he succeeded at managing his various research projects during his PhD. Finally, Mathieu is a very nice person to work with: he consistently provides constructive criticism and proposes new ideas, he also interacts easily with other researchers to create new collaborations.

In passing, let me mention that Mathieu's work has been presented at major CS conferences: *Symposium on Computational Geometry* (flagship conference, 2 articles), *International Conference on Machine Learning* (flagship conference, 1 article and 1 submission under review), *Symposium on Geometry Processing* (1 article). Meanwhile, it has led to publications in prestigious journals in applied mathematics and computer science, e.g. *Foundations of Computational Mathematics* (1 article) and *Journal of Machine Learning Research* (1 article). This is not only an excellent publication rate for our area, but also, the breadth and variety of the topics covered by these publications are remarkable.

In conclusion, I dare say that the three years spent with Mathieu to do research together during his PhD have been extremely exciting and fruitful. He definitely has the qualities and behavior to become a prominent researcher in our field, at the interface between TDA, statistics and machine learning. I therefore recommend him wholeheartedly and without any restriction for the Prix de la Chancellerie.

Sincerely,

Steve Oudot