Statistical Aspects of Topologigal Data Analysis
Internship proposal

Duration: 5/6 months.

Topic: computational geometry, topological data analysis.

Institution and Lab: GEOMETRICA team at INRIA Saclay (Turing building on the Polytechnique Campus).
Équipe GEOMETRICA, INRIA Saclay Île-de-France
1 rue Honoré d'Estienne d'Orves
Bâtiment Alan Turing
Campus de l'École Polytechnique
91120 Palaiseau
FRANCE

Advisor: Frédéric Chazal
http://geometrica.saclay.inria.fr/team/Fred.Chazal

Contact: frederic.chazal@inria.fr

General presentation: The wide availability of measurement devices and simulation tools has led to an explosion in the amount of available data, both in academia and in industry. Often these data are in the form of samples from some underlying geometric space or entity. Before such data can be effectively exploited, it needs to be processed so its underlying structures can be identified, extracted, and analyzed. In the last decade or so, computational topology was a major contributor to the understanding of geometric structures in point cloud data. In particular the emergence of Topological Persistence theory [4, 5] has provided very efficient tools to study the geometric structure of data. The goal of this internship is to design, implement and experiment new well-founded statistical approaches for topological data analysis.

Expected work: A first short period of the internship will be devoted to the learning of the basics of topological persistence and topological data analysis as well as (if needed) the basics of statistics needed to address the considered questions.
Then, we will use recent results in geometric inference and persistence topology [3, 2] and a newly introduced concepts in persistent homology [1] to study different topological properties of data sets from a statistical point of view. The goal is two fold: we expect to obtain both mathematically well-founded results and efficient implemented tools. The relevance and efficiency of the designed tools will be tested on synthetic and real data sets.

Expected knowledge and background: A background in computational geometry/topology. Some notions of C/C++ or Python would be welcome.

References

[1] P. Bubenik. *Statistical topology using persistence landscapes.* arXiv:1207.6437v1 [math.AT], 2012.

[2] F. Chazal, V. de Silva, S. Oudot, *Persistence Stability for Geometric complexes*, arXiv:1207.3885, July 2012: http://arxiv.org/abs/1207.3885

[3] F. Chazal, D. Cohen-Steiner, Q. Mérigot, *Geometric Inference for Probability Measures*, Journal on Foundation of Computational Mathematics, 11, 6, 2011.

[4] H. Edelsbrunner and J. Harer. *Persistent homology — a survey.* Surveys on Discrete and Computational Geometry. Twenty Years Later, 257-282, eds. J. E. Goodman, J. Pach and R. Pollack, Contemporary Mathematics 453, Amer. Math. Soc., Providence, Rhode Island, 2008.

[5] H. Edelsbrunner and J. L. Harer. *Computational Topology. An Introduction*. Amer. Math. Soc., Providence, Rhode Island, 2010.