

MATHEMATICAL AND COMPUTATIONAL LINGUISTICS PROJECT N.1

PERSISTENT HOMOLOGY OF SYNTACTIC PARAMETERS

CS101/MA191 CLASS WINTER 2015 TAUGHT BY MATILDE MARCOLLI

1. TOPOLOGICAL DATA ANALYSIS

In recent years, a new approach to data analysis has been developed, based on topological methods. The basic idea is to understand structures in a (large) set of data in a (high-dimensional) space, by associating to it a simplicial topological space and studying its topology. The starting point is a set of data with a proximity parameter (such as a distance function). The simplicial complex (Vietoris-Rips complex) is constructed by taking the set of data as the vertex set and assigning a k -dimensional face (k -simplex) to a $k + 1$ -tuple of data $\{x_0, \dots, x_k\}$ iff the distances satisfy $d(x_i, x_j) \leq \epsilon$ for all $0 \leq i, j \leq k$. Other versions of simplicial complexes associated to sets of data are described in [1]. General introductions to topological data analysis can be found in [1], [2], [3], [6]. Software packages are available at [5].

2. SYNTACTIC PARAMETERS

We can consider as set of data the syntactic parameters of the world languages, as collected in the database [4]. The purpose of the project is to subject this set of data to topological analysis and compute some of the topological invariants associated to it, as a way to study its structure.

REFERENCES

- [1] G. Carlsson, *Topology and Data*, Bulletin of the American Mathematical Society, Vol. 46 (2009) N.2, 255–308.
- [2] H. Edelsbrunner, J.L. Harer, *Computational Topology: An Introduction*, American Mathematical Society, 2010.
- [3] R. Ghrist, *Elementary Applied Topology*, CreateSpace, 2014.
- [4] SSWL Database of Syntactic Parameters:
<http://sswl.railsplayground.net/>
- [5] Perseus Software Package for Persistent Cohomology
<http://www.sas.upenn.edu/~vnanda/perseus/>
- [6] S. Weinberger, *What is ... Persistent Homology?* AMS Notices, Vol.58 (2011) N.1, 36–39