

# APPLICATIONS OF PERSISTENT HOMOLOGIES TO THE VIRAL GENETICS

KONSTANTIN SHMELKOV

In the last decade an interesting branch of the data analysis has emerged — the topological data analysis. The main technique of this field — “persistent homologies” — uses topological invariants to analyze a large data set represented as a point cloud in a high-dimensional space. Theoretical foundations of the persistent homologies are given in [1, 2]. Recently, an interesting application to the viral genetics was published in [3].

I will briefly describe what the topology can tell about the viral evolution and why it has a clinical interest. Then I will present a simple theoretical model of the viral genome that can be analyzed directly (in particular, formulas for Betti numbers). The nature of persistent homologies suggests computational experiments. Computations over real genomes of the influenza virus and the Hepatitis B virus as well as simulated genomes will be also discussed.

## REFERENCES

- [1] H. Edelsbrunner, D. Letscher, A. Zomorodian, “Topological persistence and simplification”, *Discrete and Computational Geometry*, Vol. 28, No. 4, 511–533 (2002).
- [2] A. Zomorodian, G. Carlsson, “Computing persistent homology”, *Discrete & Computational Geometry*, Vol. 33, No. 2, 249–274 (2005).
- [3] J. M. Chan, G. Carlsson, R. Rabadan, “Topology of viral evolution”, *Proceedings of the National Academy of Sciences*, Vol. 110, No. 46, 18566–18571 (2013).

NOVOSIBIRSK STATE UNIVERSITY, PIROGOVA STR, 2, NOVOSIBIRSK, 630090, RUSSIA  
*E-mail address:* konstantin.shmelkov@polytechnique.edu