Topological data analysis for cosmology: theory and applications

Pratyush Pranav
(Centre de Recherche Astrophysique de Lyon, Ecole Normale Superieure de Lyon)
E-mail: `pratyush.pranav@ens-lyon.fr`

Cosmology and *Big data* or data analysis, of which topo-geometrical data analysis is rapidly becoming a main component, are both burgeoning and increasingly related fields at the moment. Cosmology is transitioning from a theoretical discipline towards one with increased focus on observations, resulting in a massive surge of data that demands increasingly more sophisticated methods to glean meaningful information. In a related development, geometry and topology have witnessed a tilt from purely theoretical fields towards strong focus on application. A foray into "big data" quickly brings to front two of the central statistical challenges of our times – detection and classification of structure in extremely large, high-dimensional, data sets. Among the most intriguing new approaches to this challenge is "TDA," or "topological data analysis," the primary aim of which is providing topologically informative pre-analyses of data, which serve as input to more quantitative analyses at a later stage. Algebraic and computational topology are the foundational pillars on which TDA rests.

I will present a survey on the theoretical and computational aspects of topological data analysis [1], simultaneously exploring up the application component via analyses of cosmological datasets. The dataset we will focus on is of the Cosmic Microwave Background, obtained by the recently concluded Planck mission, with a view to compare the observations with the predictions of the standard cosmological model, that predicts the initial conditions in the Universe to be that of an isotropic, homogeneous Gaussian random field [2]. At the epoch of recombination in the infant stage of the Universe, some 370,000 years after the Big Bang, matter and radiation separate for the first time, and radiation permeates freely in the Universe. This free-streaming radiation, that we observe as the Cosmic Microwave Background, encodes a treasure trove of information about the initial conditions and properties of matter distribution in the Universe [3].

The tentative outline of my presentation, in three parts, is as follows:

- A review of the theoretical background on geometry and topology consisting of Minkowski functionals, homology and its hierarchical extension persistent homology.
- A description of the main computational components for a variety of settings relevant to cosmological data sets, such as particle distributions and images in 2D and 3D. I will give a brief but in-depth account of the computational backbone, which relies on appropriate meshing of the domain, and hierarchical embedding of levelsets in filtration data structure.
- Building up on the first and the second item, I will present case studies involving the CMB and the SDSS dataset.

The above mentioned tentative structure is subject to time constraints, and the second item may be expunged for short duration of talk.

References

[1] Pratyush Pranav et. al. The topology of the cosmic web in terms of persistent Betti numbers. *Monthly Notices of the Royal Astronomical Society*, 465 (4) : 4281-4310, 2017.

[2] Pratyush Pranav et. al. Topology and Geometry of Gaussian random fields I: on Betti Numbers, Euler characteristic and Minkowski functionals. *Monthly Notices of the Royal Astronomical Society*, 485 (3): 4167-4208, 2019.

[3] Pratyush Pranav et. al. Unexpected Topology of the Temperature Fluctuations in the Cosmic Microwave Background. *Astronomy & Astrophysics*, 627: A163, 2019.