

This article has been accepted for publication in Monthly Notices of the Royal Astronomical Society. ©: 2023 The Authors. Published by Oxford University Press on behalf of the Royal Astronomical Society. All rights reserved.

Link to article on OUP website:

<https://academic.oup.com/mnras/article/522/2/2697/7129029?login=true>

Wasserstein distance as a new tool for discriminating cosmologies through the topology of large-scale structure

Maksym Tsizh ^{1,2}★ Vitalii Tymchyshyn^{3,4} and Franco Vazza ^{2,5,6}

¹*Astronomical Observatory of Ivan Franko National University of Lviv, Kyryla i Methodia str 8, UA-79005 Lviv, Ukraine*

²*Dipartimento di Fisica e Astronomia, Università di Bologna, Via Gobetti 92/3, I-40121 Bologna, Italy*

³*Synergetics department, Bogolyubov Institute for Theoretical Physics, Metrolohichna 14-b, UA-02000 Kyiv, Ukraine*

⁴*Data Science Center, Kyiv Academic University, UA-03142 Kyiv, Ukraine*

⁵*Institute of Radioastronomy–INAF, Via Gobetti 101, I-40129 Bologna, Italy*

⁶*Hamburger Sternwarte, Gojenbergsweg 112, D-21029 Hamburg, Germany*

Accepted 2023 April 12. Received 2023 April 10; in original form 2023 January 24

ABSTRACT

In this work, we test Wasserstein distance in conjunction with persistent homology as a tool for discriminating large-scale structures of simulated universes with different values of σ_8 cosmological parameter (present root-mean-square matter fluctuation averaged over a sphere of radius 8 Mpc comoving). The Wasserstein distance (a.k.a. the pair-matching distance) was proposed to measure the difference between two networks in terms of persistent homology. The advantage of this approach consists in its non-parametric way of probing the topology of the cosmic web, in contrast to graph-theoretical approach depending on linking length. By treating the haloes of the cosmic web as points in a point cloud, we calculate persistent homologies, build persistence (birth–death) diagrams, and evaluate Wasserstein distance between them. The latter showed itself as a convenient tool to compare simulated cosmic webs. We show that one can discern two cosmic webs (simulated or real) with different σ_8 parameter. It turns out that Wasserstein distance’s discrimination ability depends on redshift z , as well as on the dimensionality of considered homology features. We find that the highest discriminating power this tool obtains is at $z = 2$ snapshots, among the considered $z = 2, 1$, and 0.1 ones.

Key words: cosmological parameters – large-scale structure of the Universe.

1 INTRODUCTION

The quantitative exploration of the large-scale distribution of matter as a complex web has been consolidated in this century. The notion of the cosmic web, traced by matter haloes and the galaxies they contain, has emerged and confidently established in the humankind’s view on the Universe, dubbing the complex large-scale distribution of matter.

The cosmic web is often observed using galaxies as pinpoint for the underlying matter distribution, thanks to large optical surveys such as Sloan Digital Sky Survey (SDSS; Tegmark et al. 2004), 2 Micron All-Sky Survey (Huchra et al. 2012), and VIMOS Public Extragalactic Redshift Survey (Guzzo et al. 2014). On the other hand, only in recent years, signatures from the diffuse gas in filaments have been reported with stacking techniques in the X-ray (Tanimura et al. 2020), in the microwaves through the Sunyaev–Zeldovich effect (de Graaff et al. 2019), and at radio frequencies (Vernstrom et al. 2021).

The many potential large-scale correlations between observable quantities and the intrinsic topological and morphological properties of the cosmic web call for a robust geometrical and topological characterization of its global network. Many numerical algorithms have been developed over the years in order to capture and describe the complex hierarchy of structures in the matter web of

cosmological simulations (e.g. Cautun et al. 2014; Libeskind et al. 2018, for reviews).

The complex networks approach has also found its application in large-scale analysis. It treats haloes or galaxies as the vertices of a complex network (graph) and exploits the network characteristics (metrics) to shed light on the nature of the cosmic web. For example, it can be used to determine the type of structure to which a halo belongs (Tsizh et al. 2020), or it can relate network metrics with observable quantities of galaxies of the cosmic web (de Regt et al. 2018). Network analysis also makes it possible to quantitatively compare the degree of self-organization and complexity of the architectures of entirely dissimilar systems, like the cosmic web and the human brain (e.g. Vazza & Feletti 2020). The neighbouring method of analysis, typical for graph-(network-)-like data, is topological data analysis that started to conquer its place in the discussed field. In particular, its subsection called *persistent homology* will be of interest to this work.

Persistent homology is characterized by a set of suitable tools to process the structures of the cosmic web, which has been recently explored by a few works. The first introduction of persistent homology into cosmic web science was probably made in 2011 by van de Weygaert et al. (2011). During the next decade, this group did systematic work studying persistence (birth–death) diagrams, Betti numbers and curves, persistence curves, and other instruments for the cosmic web. The same group has recently studied the evolution of persistence of homologies in Lambda cold dark matter (Λ CDM) simulations (e.g. Wilding et al. 2021). The follow-up work discovered the multiscale nature of the simulated cosmic web by estimating the

* E-mail: maksym.tsizh@lnu.edu.ua

persistence of topological features (holes) of different dimensions at different scales (Bermejo et al. 2022). While the aforementioned works explored the results of the cosmological simulations, the real data of SDSS catalogue were probed with this approach in Kimura & Imai (2017), where the authors introduced the idea of measuring distances between persistent diagrams of large-scale structures. The persistent voids and filaments search in SDSS was conducted in Xu et al. (2019), while the analysis of aperture masses obtained from cosmic shear as a result of gravitational lensing has been performed with persistent homology to constrain cosmological parameters in recent works by Heydenreich, Brück & Harnois-Déraps (2021) and Heydenreich et al. (2022).

The persistent homology approach can be also applied to continuous reconstructions of the large-scale structure data, as was shown in Cisewski et al. (2014), where the authors analysed persistent homology of the H I density field in the intergalactic medium, using the Lyman-alpha forest data. Similarly, the continuous field of reionization bubbles has also become the object of persistent homologies studies in Elbers & van de Weygaert (2019, 2022). Finally, the simulation of the interstellar medium together with its magnetic field was subjected to topological data analysis in Makarenko et al. (2018).

Another well-known way of exploring the large-scale structure of the Universe is by comparing the distribution of perturbations to the Gaussian random field. Topology data analysis and persistent homology in particular, manifest themselves as suitable instruments for this problem, as can be seen from Feldbrugge et al. (2019), Biagetti, Cole & Shiu (2021), Pranav (2021), and Biagetti et al. (2022). Gaussianity of cosmic microwave background radiation was studied with persistence diagrams in Cole & Shiu (2018).

One of the most prominent applications of persistent homology in recent years is, in our opinion, the work by Cisewski-Kehe et al. (2022), in which the authors directly develop the idea of discriminating power of persistence diagrams for cosmological models. They analysed several metrics for computing the distance between persistence diagrams to check whether it is possible to tell apart the statistical difference between the persistence homologies of cold and warm dark matter universes. This, in principle, coincides with the idea of this work, which is to utilize the Wasserstein distance for the analysis of the cosmic web and to monitor its sensitivity to variations of the σ_8 cosmological parameter, i.e. the present rms matter fluctuation averaged over a sphere of radius 8 Mpc comoving (see also Biagetti et al. 2021, where the possibility of persistence homology methods to resolve between the impact of different σ_8 and non-Gaussianity parameter on cosmic web was shown).

The graph-theoretical approach, known also as the complex network approach, discovers many aspects of the cosmic web (see e.g. an overview in Tsizh et al. 2020), but it still poses a significant disadvantage when applied to graph-like data. It consists in the critical dependence of the constructed graph on linking length distance, a distance under which the two haloes or galaxies are considered connected. The persistent homology is one of the possible non-parametric ways to explore the data, it liberates the explorer from disadvantage of casting pre-defined scale onto the analysis.

We present our work as follows. In Section 2, we present the core idea of persistence homology and convey the definition of Wasserstein distance two between persistence diagrams. In Section 3, we describe the simulation we used for analysis and we formulate the hypothesis to test. In Section 4, we present our main results. In Section 5, we discuss limitations of our findings and the possibility of applying the new tool to the observational catalogues of galaxies. In the last section, Section 6, we summarize the paper.

2 PERSISTENT HOMOLOGY

A number of computational problems in theoretical physics can be reduced to comparing points or clouds of points, be these points measured or calculated positions of real physical objects (e.g. stars, galaxies, gas molecules) or even states of a certain system in its configuration space. It turns out, there is no straightforward notion of the distance between two point clouds: considering each individual point leads to ‘informational overflow’, ambiguity, and in most cases will be too sensitive to small changes in the positions of points. To solve the problem, one can avoid considering individual points, but extract instead certain generalized information about their distribution that will comprise a manageable amount of numbers and then look for some notion of distance for these sets of numbers.

One of the possible approaches is to endow each point with a sphere of radius r centred at the point and consider the union of all these balls as a manifold. This manifold will have certain topological properties that can be numerically represented and used further. As such, one uses Betti numbers – ranks of appropriate homology groups of the given manifold. For simplicity, one can think of 0-Betti number as a number of connected components, 1-Betti number as a number of one-dimensional or ‘circular’ holes, 2-Betti number as a number of two-dimensional ‘voids’ or ‘cavities’, and so on. This notion is quite convenient for the problem under consideration as we can associate (unfortunately, a bit indirectly) Betti numbers with cosmological structures: 0th with clusters, 1th with cycles and tunnels formed by cosmological filaments, and 2th with cosmological voids formed by cosmic sheets. Moreover, representation through Betti numbers is quite stable with respect to translation, rotation, and small variation of points’ positions which is a desired property.¹

However, the choice of the radius r to use is debatable. As a solution, one can consider all possible values of r and track topological features that generate Betti numbers as the manifold changes with the change of r . Hence, the name ‘persistent homology’: each topological feature persists within a certain range of the parameter r and Betti numbers change when features are ‘born’ or ‘die’. The result can be represented as a plot r versus Betti numbers, or we can map ‘birth’ and ‘death’ of each individual topological feature in the form of a barcode diagram or persistence diagram. Interestingly, there is a notion of distance between persistence diagrams in a strict mathematical sense called ‘Wasserstein distance’ and its partial case ‘bottleneck distance’.

2.1 Wasserstein and bottleneck distances

In probability theory and computational topology there is a widely used notion of p -Wasserstein distance (Edelsbrunner & Harer 2010)

$$W_p(X; Y) = \left(\inf_{\varphi: X \rightarrow Y} \sum_{x \in X} \|x - \varphi(x)\|_\infty^p \right)^{1/p}, \quad (1)$$

that provides distance between two multisets, X and Y (sets that allow multiple instances of any element). The idea of this distance is the

¹In the code, we extract topological features with so-called α -complex without any restriction on the parameter α . One may be concerned that it is defined in terms of simplices and Delaunay triangulation but not intersecting balls. Though, without any restriction on α it is equivalent to the Čech complex (but much smaller that is beneficial for calculation) and Čech complex in its turn is homotopy equivalent to the ‘intersecting spheres’ point of view by the classic ‘Nerve theorem’ (Alexandroff 1928). Thus both views are equivalent, but ‘intersecting spheres’ are easier to grasp.

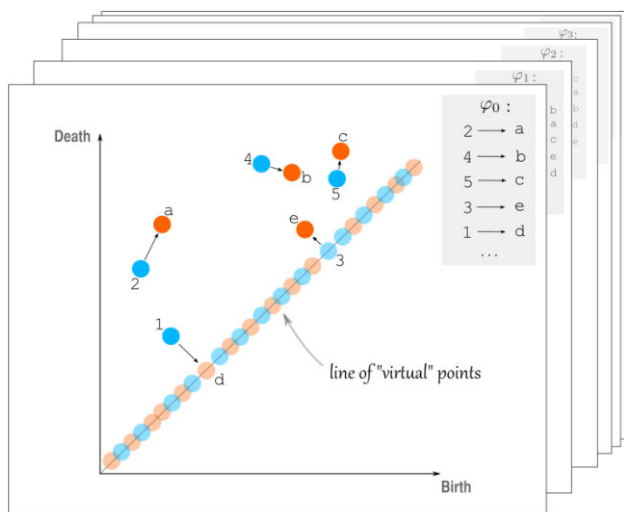


Figure 1. Maps φ between persistence intervals as in (1). Blue and red dots represent persistence intervals calculated from two different point clouds, their XY coordinates represent the endpoints (‘births’ and ‘deaths’) of corresponding persistence intervals.

following: first, we match elements of X and Y one-to-one (bijection φ) in a certain way (suppose it can be performed for now). For each pair of elements, $x \in X$ and $\varphi(x) = y \in Y$, we can calculate how much do they differ (the cost function) $\|x - \varphi(x)\|_\infty$ that is basically L_∞ norm. Adding up the p th degrees $\|\cdot\|_\infty^p$, we get a notion of the difference between the whole multisets X and Y under the matching $\varphi: X \rightarrow Y$. Taking the infimum over all possible bijections φ , we get the difference between multisets X and Y under the best matching possible, effectively removing φ from further consideration. Now, taking the root of p -th order transforms the whole expressions into a well-defined distance that satisfies all axioms of metric (Figalli & Glaudo 2021).

Let us consider this distance with respect to our problem. As we tracked how Betti numbers change with respect to radius r , the topological features were ‘born’ and ‘died’. Thus, we got a persistence interval $[r_b; r_d]$ for each feature (can be represented as a point on a plane). X and Y will now represent sets of such intervals calculated on different point clouds. The L_∞ norm for two such intervals is defined as

$$\| [r_b^A; r_d^A] - [r_b^B; r_d^B] \|_\infty = \max(|r_b^A - r_b^B|, |r_d^A - r_d^B|),$$

where $r_b^A, r_d^A, r_b^B, r_d^B$ are ‘birth’ and ‘death’ radii. Plugging it into (1), we get the desired expression.

One more adjustment in the procedure is necessary since the persistence diagrams consist of finitely many points above the diagonal that can vary in number depending on the initial point cloud they were generated from. To this finite multiset, we add the infinitely many points on the diagonal, each with infinite multiplicity (Edelsbrunner & Harer 2010) (‘virtual points’) and allow matching with them. These extra points are not essential to the diagram, but their presence allows us to find bijection φ even in cases there are different numbers of persistence intervals in X and Y as shown in Fig. 1. One may note that taking infimum over all possible φ , we will always end up in a situation when most of the diagonal points are mapped to the diagonal points with the same coordinates, thus adding 0 (i.e. smallest possible $\|x - \varphi(x)\|_\infty^p$) to the total cost function. Only the ones matched with the off-diagonal points will add to the total cost.

The bottleneck distance is the Wasserstein distance, with parameter $p \rightarrow \infty$. Finding the appropriate limit, it can be shown that (Edelsbrunner & Harer 2010)

$$W_\infty(X; Y) = \inf_{\varphi: X \rightarrow Y} \sup_{x \in X} \|x - \varphi(x)\|_\infty. \quad (2)$$

The precursor of the modern notion of bottleneck distance was probably first introduced by Patrizio Frosini in 1990 (Frosini 1990).

3 COSMOLOGICAL SIMULATIONS

For this work, we analysed a set of recent cosmological simulations produced with the magnetohydrodynamical code ENZO,² applied to a suite of concordance Λ CDM simulations of eight different cosmic volumes, for six different values of the σ_8 parameter. We used here a simple uniform resolution grid approach, sampling a comoving volume of 42.5^3 Mpc^3 with 512^3 cells (yielding a constant spacial resolution of 83.3 kpc per cell) and with 512^3 dark matter particles (yielding a fixed mass resolution $m_{\text{dm}} = 6.19 \times 10^7 M_\odot$).

These runs are part of a larger project, aiming at investigating the origin of cosmic magnetism through multiple resimulations of magnetic field seeding scenarios (e.g. Vazza et al. 2021, and references therein). However, unlike what we did in most of the other projects along this line of research, here we fixed the initial magnetic field (assumed to have a primordial origin) and studied how the properties of simulated magnetic fields change with increasing initial amplitude of density perturbation (σ_8 parameter) and are also affected by cosmic variance.

Each simulation starts at $z = 40$ from a ‘primordial’ uniform volume-filling comoving magnetic field $B_0 = 0.1 \text{ nG}$ for each magnetic field component.

The simulation also includes the effect of radiative (equilibrium) cooling on baryon gas, assuming for simplicity a primordial chemical composition, and no additional sources of feedback. While these effects are not of primary importance for the study of homology presented here (which could have been done just using N -body simulations of dark matter, as usually done in the literature, the adoption of non-gravitational physics adds realism to the properties of the simulated gas network of the cosmic web, and it allows us to produce mock observables for these runs (subject of future works).

The cosmological parameters in this suite of simulations are kept constant to the reference values of a flat Λ CDM cosmological model, with $H_0 = 67.8 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_M = 0.308$, $\Omega_\Lambda = 0.692$, $\Omega_b = 0.0468$, and $\sigma_8 = 0.815$ (Planck Collaboration XIII 2016).

With this set-up, we generated a small suite of 48 simulations, evolved from $z = 40$ to $z = 0.0$, for eight random variations of the initial phases of the matter and velocity distributions in the initial conditions (to produce eight independent random realizations of the same cosmology and gauge the effect of cosmic variance) as well as six different simulations with uniformly increasing σ_8 , from 0.5 to 1.0.

We notice that compared to the recent work by Bermejo et al. (2022), who first analysed the simulated cosmic web using persistent homology, our suite of simulations investigates a much smaller (by a factor ~ 800) cosmic volume, while on the other hand, it provides a ~ 10 better mass resolution for haloes, and it also allows us to monitor the effect of σ_8 and cosmic variance through the comparison of different resimulations.

²enzo-project.org

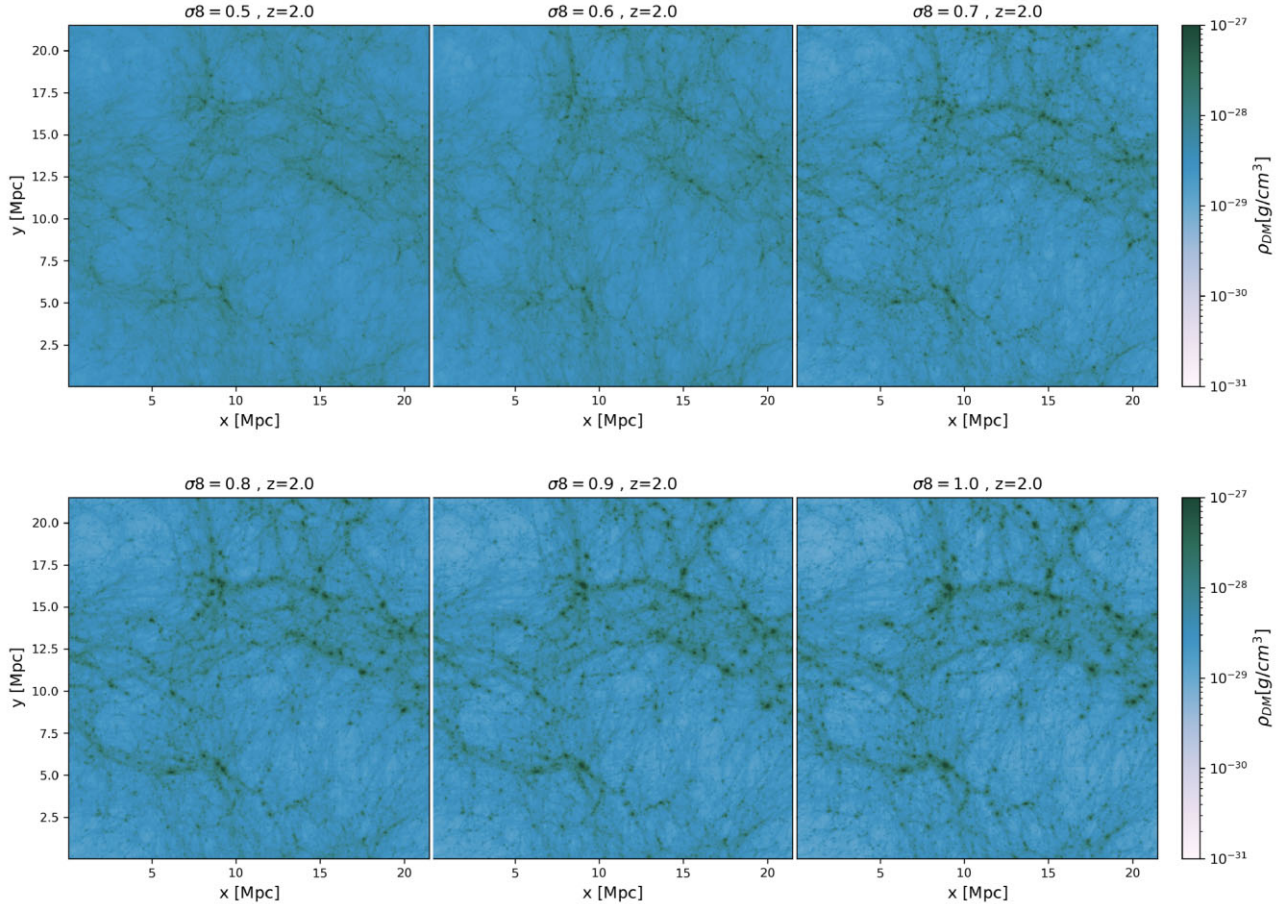


Figure 2. Projected maps of the average dark matter density along the full 42.5 Mpc line of sight, for all variations of σ_8 for 1 of our 10 randomly extracted initial conditions, for the epoch of $z = 2.0$.

Figs 2 and 3 give the visual example of the increased level of clustering of the dark matter component in one of our random seed extractions of the initial conditions, for the different σ_8 and at the early epoch of $z = 2.0$, and towards the end of the simulation ($z = 0.1$).

The masses and positions of self-gravitating haloes in all simulations are computed using the parallel friends-of-friends (FOF) algorithm in ENZO, imposing for all runs a linking length of 0.05 cells and a minimum number of 50 dark matter particles for each halo, and computing their M_{200} masses.

The dark matter mass distribution of haloes in all our boxes and for all eight random variations of initial seeds is shown in Fig. 4. As for the projected masses of dark matter distribution, the effect of a decreasing σ_8 is seen, as expected, in a progressively decreased normalization of the measured mass function, which also leads to slower growth time of the most massive haloes in the boxes.

On purpose, we do not present a detailed comparison with the theoretical expectations for the mass functions under the same cosmological models, because there are small, but not entirely negligible effects related to non-gravitational physics, which can affect the timing of halo formation as a function of σ_8 , especially for the lowest values of it.

Indeed, while all adopted non-gravitational effects are known to lead to negligible differences in the abundance of haloes at low redshift (at least, for the range of baryonic physics and magnetic effects considered here), it is non-obvious to assess their impact at

high redshift, and in the regime of initially low-density perturbations. On one hand, a uniform magnetic field level can slow down the collapse of haloes, by providing extra pressure to the gas (e.g. Dolag, Bartelmann & Lesch 1999; Kahnishvili et al. 2013), while on the other the impact or radiative gas cooling is that of accelerating the collapse of haloes, leading to an increase up to a factor two with respect to non-radiative simulations at high redshift (e.g. Cui et al. 2012).

Our suite of simulations has not been designed uniquely for this project, but also to compare with real radio observations of cosmic magnetism (Carretti et al. 2023), hence taking into account the presence of non-gravitational physics and its possible (secondary) effect on clustering. In any case, it is interesting to study how the global effect of gravity and baryonic physics can alter the measured distribution of haloes as a function of redshift and σ_8 , which is the case that real observations have to face.

More ad hoc tests, in which single dependencies related to the input σ_8 can be tested and isolated through persistent homology will be considered in future work.

4 RESULTS

First, we can make a qualitative comparison of our studies to others. Persistent homology was already applied to large-scale data. We can compare our (Fig. 5) persistence (birth-persistence in this case) diagrams at different z to those, for example in Biagetti et al. (2021);

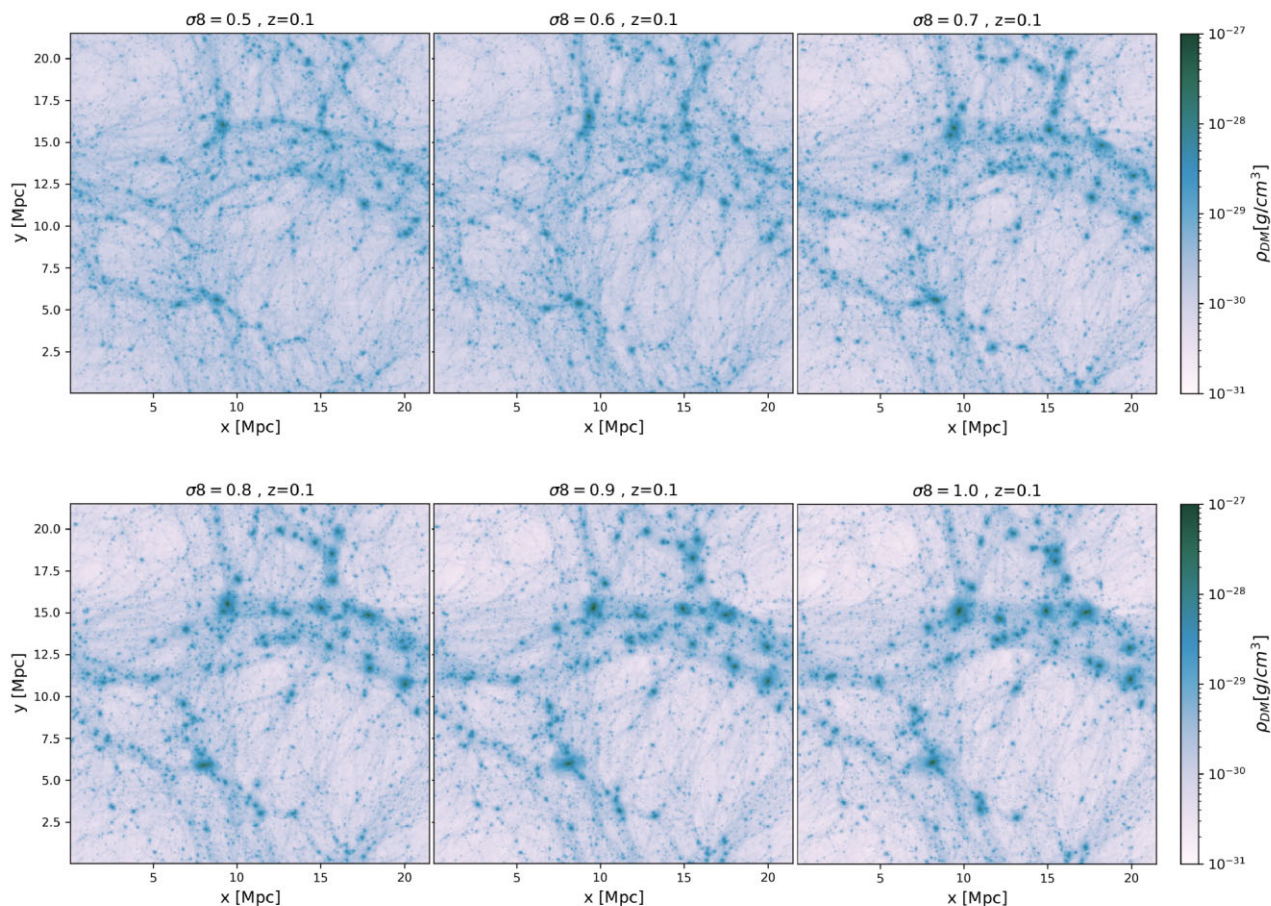


Figure 3. Projected maps of average dark matter density along the full 42.5 Mpc line of sight, for the same model used in Fig. 2, for the final epoch of $z = 0.1$.

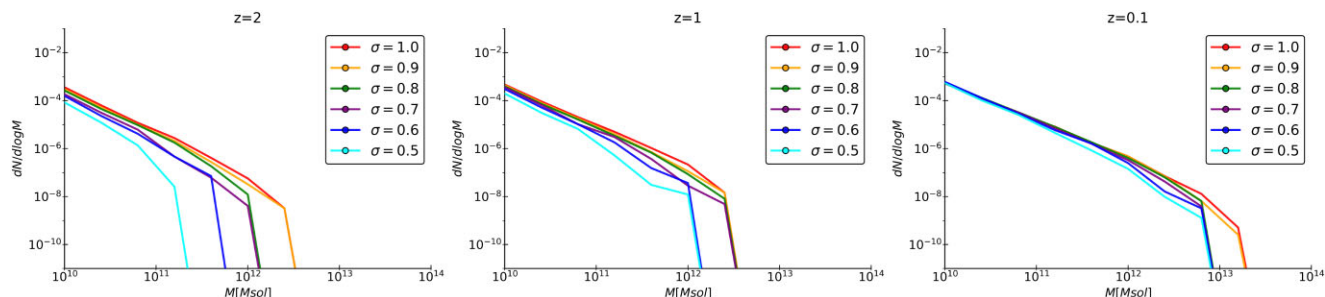


Figure 4. Halo mass functions (normalized to a comoving volume of 1 Mpc^3) for all haloes in our suite of simulations, for three different redshifts and for the six different simulated values of σ_8 (solid lines).

fig. 6 in there). One can notice that these images, in principle, have common tendencies. These are

(i) persistence of the features on smaller scales at lower z . The points that correspond to $z = 0.1$ are (on average) on the left to those which correspond to $z = 1$, which are on the left of those that correspond to $z = 2$.

(ii) 2-homologies have wider and less steep distribution and on average, they are born at a slightly larger radius.

These common tendencies show qualitative agreement between our study to those done before.

Now, we will graphically present and elaborate on the numerical results. We have computed three types of defined above Wasser-

stein distances between each pair of simulated universes: the 1-Wasserstein, 2-Wasserstein, and ∞ -Wasserstein (or the ‘bottleneck’) distance. The distances are measured between the ‘birth–death’ diagrams of each pair. We compute distances for zero-dimensional, one-dimensional, and two-dimensional homology features (0-, 1-, and 2-homologies). While 1- and 2-homologies can be treated as cycles of filaments and voids of the cosmic web, respectively, 0-homologies are nothing else but connected components that start as separate points at $r = 0$.

The distances, unless mentioned otherwise, are given in a range between 0 and 1, this is scaled values, where 1 corresponds to the length of the edge of a simulation cube, 42.5 Mpc. We are interested in the dependence of these distances on the difference in σ_8 parameter

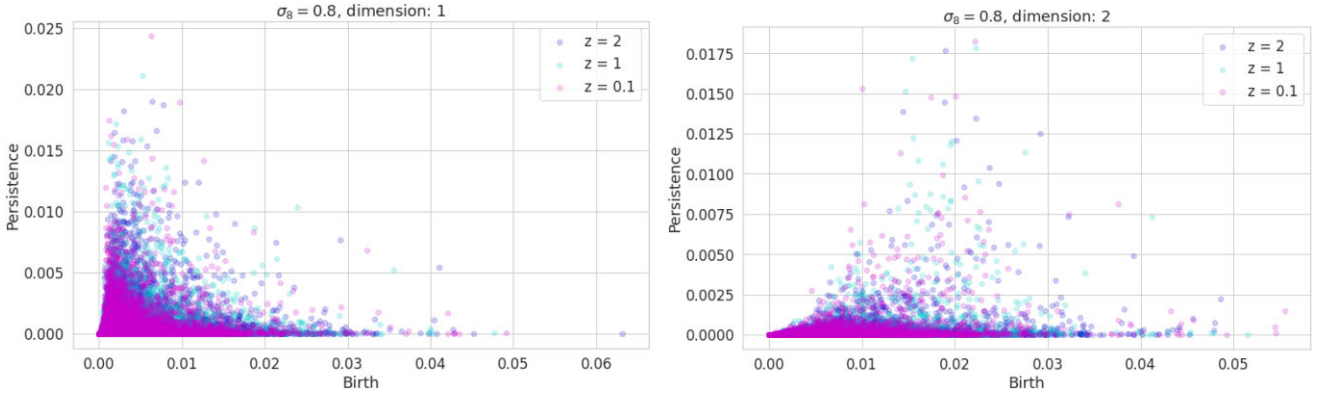


Figure 5. Persistence-birth diagrams for 1- (left) and 2- (right) homologies for universes with $\sigma_8 = 8$ at different redshifts.

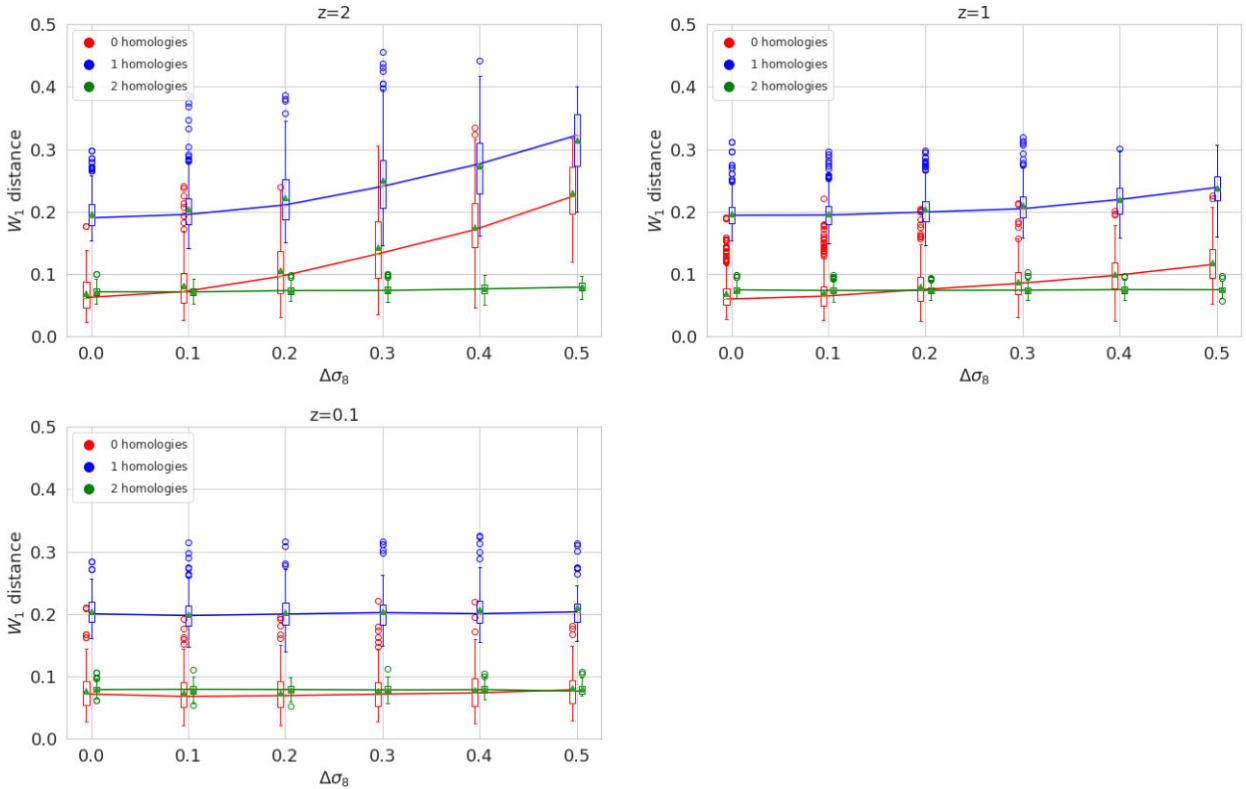
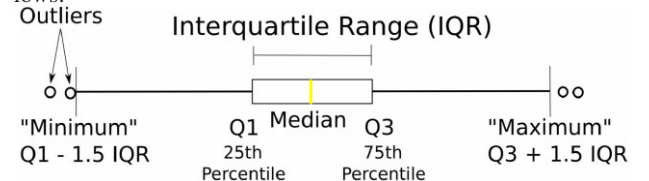


Figure 6. Boxplots of W_1 distances at different z . Each figure contains three sets of data for homologies of different dimensions (colour-coded). Solid lines represent the general trend for median values. Additionally, mean values are shown as triangles. Please note, data for 0- and 2-homologies were shifted a bit to the left and right to avoid boxplots overlapping.

between simulated universes, $\Delta\sigma_8$. To take into account the impact of cosmic variance on our results, we have eight simulations for each value of σ_8 , yielding $8 \times 6 \times (8 \times 6 - 1)/2 = 1128$ pairs of point clouds, and, respectively distances calculated. We then average values of distances for each $\Delta\sigma_8$ value. The obvious disadvantage of such a way of averaging is, that we have a different number of points to the average for each value of $\Delta\sigma_8$, as the smaller $\Delta\sigma_8$ is, the larger number of pairs exist with such difference. Indeed, we have $6 \times 8 \times (8 - 1) = 336$ points at $\Delta\sigma_8 = 0$ and only $8 \times 8 = 64$ at $\Delta\sigma_8 = 0.5$. However, we have to reconcile with such a feature and rely on a large number of points even for $\Delta\sigma_8 = 0.5$.

Results for different values of z are represented in Fig. 6. The X-axis shows the difference $\Delta\sigma_8$ between two simula-

tions. The computed Wasserstein distances between persistent diagrams for different simulations can be treated in this context as a realization of a random variable (e.g. different seeds contribute to the randomness of halo distribution). Thus, we represent results in form of ‘boxplots’ – a well-known type of diagram in statistics. Its ‘anatomy’ can be represented as follows:



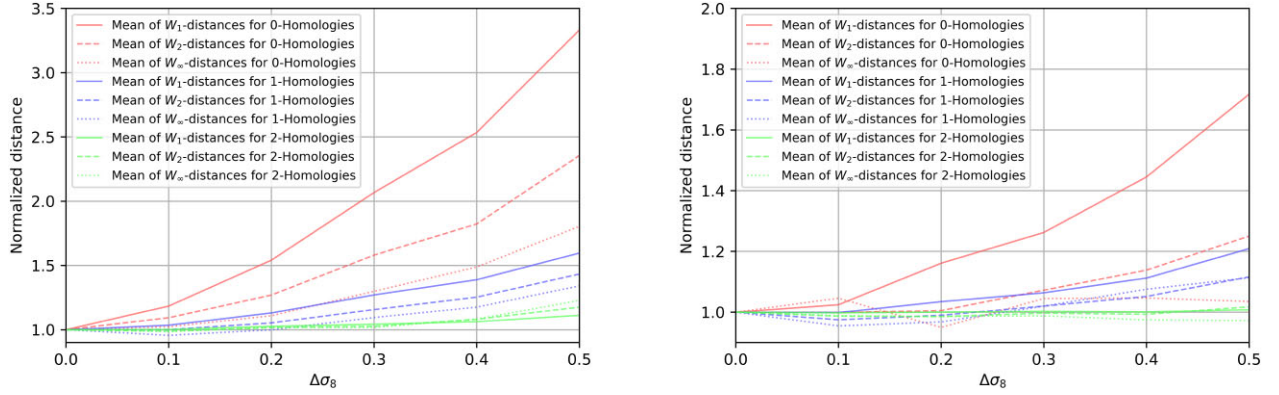


Figure 7. Comparison of normalized distances changes with change of $\Delta\sigma_8$ at $z = 2$ (left-hand panel) and $z = 1$ (right-hand panel). Note, that y-axis scale in panels is different.

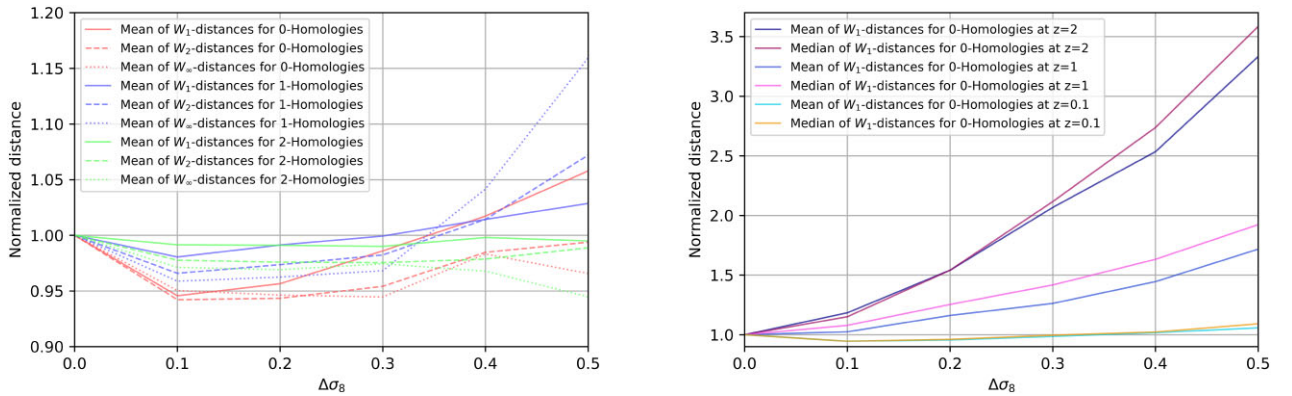


Figure 8. Comparison of normalized distances changes with change of $\Delta\sigma_8$ at $z = 0.1$ (left-hand panel). Comparison of normalized W_1 distances at different z (right-hand panel).

Data points are sorted in ascending order and split into four quartiles (Q1, ..., Q4), each containing 25 per cent of the data points. The median shows where Q2 ends and Q3 starts. Calculations were performed on persistence diagrams of different order homologies separately (blue, green, and red colours for 0-, 1-, and 2-homologies, respectively), thus three curves in each plot.

Among all analysed distances, 1-Wasserstein distances have the higher contrast between $\Delta\sigma_8 = 0$ and $\Delta\sigma_8 = 0.5$ for $z = 2$ and $z = 1$, this is why we select to show it on this figure. For $z = 0.1$, ∞ -Wasserstein distance has negligibly better contrast. Note, that each of the dimensions of homologies occupies its own niche of distances with the largest ones for 1-homologies.

In Figs 7 and 8, one can find the comparison of different variants of Wasserstein distances in distinguishing power between universes with varying σ_8 in normalized values. In these plots, each point of the line (which shows the averaged distance) is divided by its value at $\Delta\sigma_8 = 0$. In Fig. 7 and the left-hand panel of Fig. 8, the red lines represent the 0-homologies, the blue lines represent 1-homologies, and the green lines correspond to 2-homologies. Different styles of lines correspond to different types of distances (solid line to 1-Wasserstein, line-dotted to 2-Wasserstein, and dotted lines to ∞ -Wasserstein distance). On the right-hand panel of Fig. 8, one will find a comparison of distinguishing power of averages and medians of W_1 distance at different z . Medians might work slightly better in some cases.

Let us deeper analyse these two sets of graphs. The first and obvious conclusion is that on average, Wasserstein distances reflect the

size of the difference in σ_8 parameter between simulated universes. However, the distinguishing power of persistent homology decreases as redshift approaches zero. It appears that the networks of haloes are better at breaking the degeneracy of $\Delta\sigma_8$ in the earlier universes. The second conclusion is, that zero-dimensional homologies work better than 1-dimensional homologies, which, in turn, work better than two-dimensional ones. Persistent loops of filaments do a better job than persistent voids and sheets of matter, and comparing the persistence of separated components is the best of all. The averages and medians of distances for 0-homologies at $z = 2$ increase their value by more than 3.4, compared to 1.7 for 1-homologies and 1.2 for 2-homologies. Meantime, at $z = 0.1$ distances between universes with contrasting σ_8 become almost negligible. We can also note, that dissimilarities between different types of Wasserstein distances are minor, although they are still present.

At the same time, the large width of the distribution of distances does not allow us to build a rigid statistical test, which would be able to set apart simulated universes with close values of σ_8 . Indeed, as can be read from Fig. 9, even for the best case of 1-Wasserstein distance of 0- and 1-homologies at $z = 2$, the probability density functions of data points moderately overlap each other, and the situation worsens for low redshift data ($z = 0.1$). In these graphs, we estimate the particle's distribution function with the kernel density estimation method. As a kernel, we used a Gaussian kernel with the width chosen by Silverman's rule.

We have additionally checked, whether information about the differences is hidden in the 'noise' of the large population of haloes.

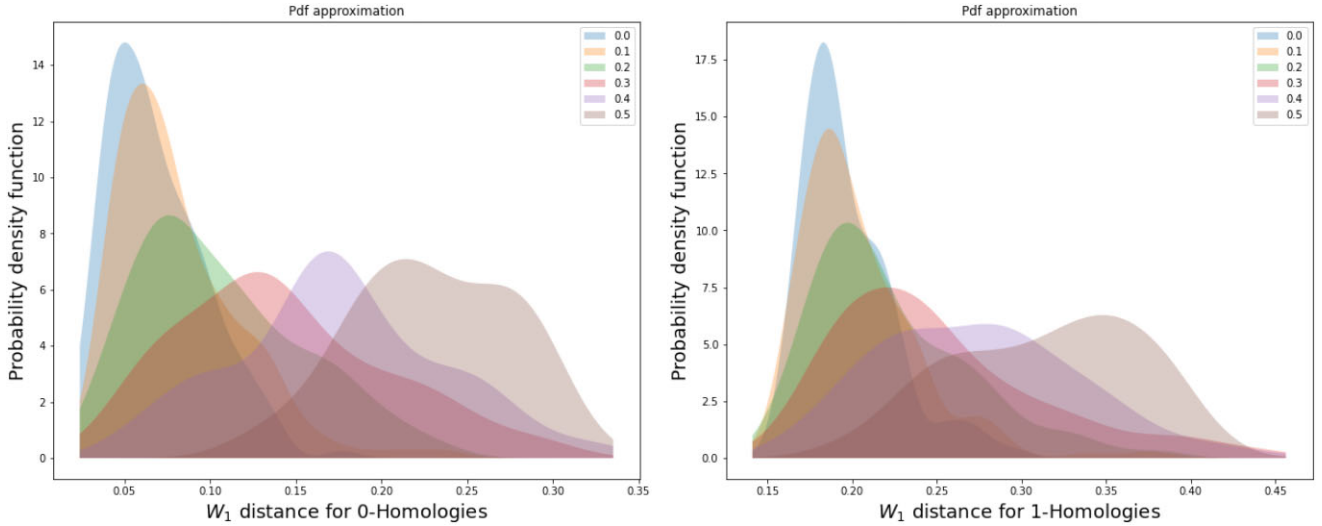


Figure 9. Approximation of probability density functions for Wasserstein distances between simulated universes at $z = 2$ for 0-homologies (on the left) and 1-homologies (on the right).

We have calculated the same distances for ‘bootstrapped’ data, where part of the haloes was randomly deleted, averaging the overall result. As can be seen for Fig. 10, bootstrapping did not improve the results. Likewise, the selective filtering out of the least or of the most massive haloes did not produce a significant improvement of our results, and in Fig. 11, one can find and compare the averages of distances in populations, where only 10 percent of heaviest and 10 percent of lightest haloes are left. The case of 0-homologies mass filtering at $z = 1$, where there is a considerable difference is rather an exception. This overall is a promising result: when applying this instrument to an even incomplete catalogue of real galaxies, it is still possible to reconstruct the underlying topology of the network connecting them, in a robust enough way to constrain the value of σ_8 .

We have also tested whether the large span in a number of haloes might cause large Wasserstein distances between universes instead of a difference in σ_8 . To do this, we conduct all the same calculations on the data set, where the number of haloes in all universes was equalized by truncation to a certain limit. We tested truncation (randomly throwing out excessive haloes) at 800, 2000, and 4000 haloes.³ If the number of haloes in a universe was lower than this limit, we use bootstrapping (randomly repeatedly selecting the haloes) to sample up this universe to the limit. In neither case, the result (sensitivity of W_1 to σ_8) was not considerably improved. In most cases, it was worsened and just in some, it was the same or slightly better. From this, we can abstemiously conclude that sampling only a small part of the population of the cosmic web would not allow one to discern the nature of its homology features. All of the computations were performed with GUDHI PYTHON library (Maria et al. 2014), a specialized package for dealing with persistent homology. The plots were built with MATPLOTLIB library of PYTHON. All of our scripts and data can be found in open access.⁴ It took a moderate amount of computational time to complete all of the calculations: a day on a modern desktop PC would be enough, which is an advantage of the considered method.

³Originally, for $z = 2$ universes the span in the number of haloes is (849, 5264), for $z = 1$ it is (2373, 7121), and for $z = 0.1$ it is (6353, 10195).

⁴<https://github.com/mtsizh/bottleneck-distance-for-sigma8>

5 DISCUSSION

First of all, it shall be noted that, although persistent homology analysis requires no arbitrary choice of spacial scales, our results are still restricted by the limited spacial and mass scales sampled with our suite simulations (i.e. from 83 kpc to 42.5 Mpc comoving and from 6.1×10^7 to $\sim 10^{14} M_\odot$ for the dark matter component). However, theoretical works suggest that the self-organization of the cosmic web evolves only slowly (i.e. logarithmically) with scale (e.g. Sylos Labini, Vasilyev & Baryshev 2007; Sylos Labini 2011). The findings of Bermejo et al. (2022) confirm the logarithmic nature of this evolution also in terms of persistent homology and confirms the multiscale nature of the cosmic web. This is why we expect our results to be possibly relevant to a wide range of spacial scales of the cosmic web.

We would like to stress on the potential that persistent homology analysis has if applied to real data (either for future or even existing, surveys of galaxies). Unlike the more standard measurement of the mass function of galaxy haloes as a cosmological probe, the persistent homology requires no knowledge of the halo masses (which can be increasingly more difficult to estimate for high redshift and/or small mass haloes). Since only the three-dimensional position of galaxies, assumed to mark the location of dark matter haloes, is needed to analyse the topology of the underlying matter distribution of the cosmic web in this approach, the only requirements are related to the accuracy of galaxy positions, in real surveys. Our additional tests also showed that this procedure is also robust against the removal of the least massive (or most massive) haloes in catalogues.

As a comparison with real sky observations, we note that our analysis concerns eight independent realizations of the same cosmology (with variations of σ_8), which effectively cover a comoving volume of 85^3 Mpc^3 . This corresponds to a projected field of view with size $\approx 2.7^\circ$ at $z = 2$, $\approx 2.9^\circ$ at $z = 1$, or $\approx 13.1^\circ$ at $z = 0.1$, meaning that the area of sky which needs to be surveyed to test this technique is affordably small. Moreover, the expected spectroscopic redshift uncertainty in future Euclid surveys is $\Delta z \approx 0.001(1 + z)$ in the 0.7–1.8 redshift range (e.g. Euclid Collaboration 2021). Although for a smaller field of view, similar redshift uncertainties also apply to existing surveys (e.g. 224 arcmin² for the VIMOS survey; Scodreggio et al. 2018, and 1.7 deg² for z-COSMOS; Lilly et al. 2007, to cite

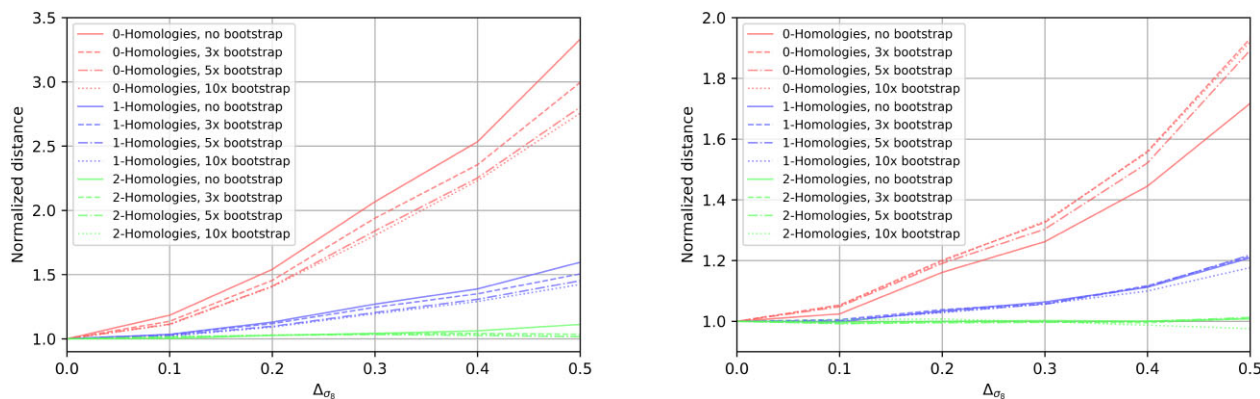


Figure 10. W_1 distances at $z = 2$ (left) and $z = 1$ (right) with bootstrapped data.

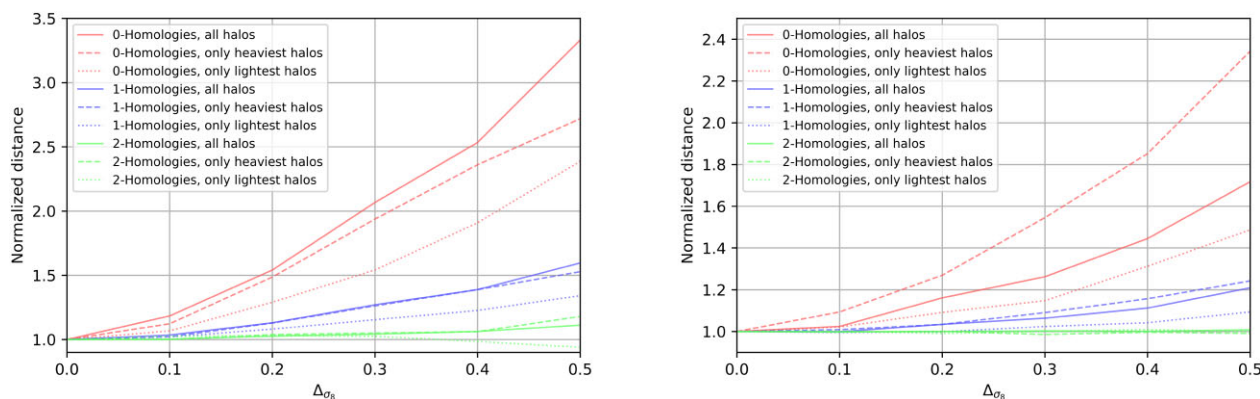


Figure 11. W_1 distances at $z=2$ (left) and $z = 1$ (right) with mass-filtered population of haloes.

a few). Significantly larger uncertainties, and restricted to a $z \leq 1$, are instead typically available in the SDSS survey (e.g. Pasquet et al. 2019). This corresponds to about 8 kpc at $z = 2$ and to about 5.5 kpc at $z = 1$, and it means that at both epochs the typical redshift uncertainty of future surveys is even smaller than the spacial resolution of the simulation we used to perform our homology analysis, and for this reason, we expect this technique to be, at least in principle, suitable to analyse the cosmic web even when realistic uncertainties in the redshift of galaxies are taken into account. Moreover, it shall be stressed that even if in this work we have mainly focused on the effect of σ_8 on the growth of factor, which in our case also carries secondary effects related to the somewhat delayed growth of haloes due to the non-linear interplay of other non-gravitational processes (such as primordial magnetic fields and reionization temperature background), in principle the same technique can be used to assess the impact of other physical parameters on the growth of structures, such as neutrino masses (e.g. Castorina et al. 2015) warm dark matter and modified gravity models (e.g. Baldi & Villaescusa-Navarro 2018), as well as other effects related to baryon physics (e.g. Shao, Anbajagane & Chang 2022).

6 CONCLUSIONS

In this work, we presented new results from the application of persistent homology to study the multiscale structure of the simulated cosmic web. Persistent homology is a powerful tool that allows us to decode the topological embedding of self-gravitating mass haloes, and instead of focusing on the topology at a single (and arbitrary)

length scale, it detects persistent topological features (measured through their Betti number) within the range of spacial scales in which they ‘born’ and ‘die’, as a result of increasing filtration. The Wasserstein distance resulting from this analysis (Section 2.1) is thus a helpful and non-arbitrary measurement of the difference of typical scale characterizing the most persistent topological features of the cosmic web. Such difference, if induced by cosmological parameters (for example, σ_8), can be traced out and utilized to distinguish between different cosmologies.

In summary, our main conclusions are as follows:

- (i) The Wasserstein distance shows the best discrimination power of σ_8 at the largest investigated redshift, $z = 2$. The discrimination becomes increasingly worse moving towards low redshift (i.e. a little worse at $z = 1$, and very low at $z = 0.1$).
- (ii) The persistence features of dimension 0 are much more sensitive to σ_8 than the features of dimension 1, which, in turn, are more sensitive than those of dimension 2. The distinguishing power decreases with the increasing dimensionality of features.
- (iii) The 1-Wasserstein distance shows the best results in comparing to 2-Wasserstein, or to the Wasserstein distance (∞ -Wasserstein), in terms of discrimination of σ_8 , but differences are really small.
- (iv) These findings are robust against bootstrapping, or the mass-filtering of the haloes used for the network reconstruction. We also find that the median value of the above distance estimates, in general, is a little more informative than the mean of distances.
- (v) The distributions of Wasserstein distances for different values of $\Delta\sigma_8$ somewhat overlap one with another. Thus, the cosmic variance prevents us from the possibility to build a rigid statistical

test of restricting the value of σ_8 by just comparing the persistence of homologies of cosmic webs.

(vi) The selected spacial and mass ranges of our study allow us to carefully suggest possible similar exploration of persistent homology of the real cosmic web, as modern galaxy catalogues possess the required size and resolution.

With all of this, we can predict, that the next steps in studies towards a deeper understanding of the topology nature of large-scale matter distribution would be exploring how other cosmological parameters shape the persistent homology of the cosmic web, as well as comparing the homologies of the observable cosmic web.

ACKNOWLEDGEMENTS

We would like to thank the Armed Forces of Ukraine for providing security to perform this work. This work has become possible only because of the resilience and courage of the Ukrainian Army. We would like to thank our anonymous reviewer for the helpful comments and suggestions, which have helped us to improve the quality of our work. MT and FV acknowledged financial support from the Horizon 2020 program under the ERC Starting Grant MAGCOW, no. 714196. We also would like to thank the Italian MUR (Ministero Università e Ricerca), the Administration of University of Bologna, and especially L. Fortunato (Physics and Astronomy Department of Unibo) for the substantial administrative support, which also made this research possible. MT was co-founded by the project of the Ministry of Education and Science of Ukraine ‘Modelling the luminosity of elements of the large-scale structure of the early Universe and the remnants of galactic supernovae and the observation of variable stars’ (state registration number 0122U001834) and also would like to thank Dr Veronika Romero and all the volunteers who help Ukrainian scientists during hard times. In this work, we used the ENZO code (<http://enzo-project.org>), the product of a collaborative effort of scientists at many universities and national laboratories. The authors gratefully acknowledged the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) for supporting this project by providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS Supercomputer JUWELS at Jülich Supercomputing Centre (JSC), under the project ‘radgalicm2’.

DATA AVAILABILITY

The data underlying this article are available in the GitHub repository at <https://github.com/mtsizh/bottleneck-distance-for-sigma8> and can be freely accessed by anyone.

REFERENCES

- Alexandroff P., 1928, *Math. Ann.*, 98, 617
 Baldi M., Villaescusa-Navarro F., 2018, *MNRAS*, 473, 3226
 Bermejo R., Wilding G., van de Weygaert R., Jones B. J. T., Vegter G., Efstathiou K., 2022, preprint ([arXiv:2206.14655](https://arxiv.org/abs/2206.14655))
 Biagetti M., Cole A., Shiu G., 2021, *J. Cosmol. Astropart. Phys.*, 2021, 061
 Biagetti M., Calles J., Castiblanco L., Cole A., Noreña J., 2022, *J. Cosmol. Astropart. Phys.*, 2022, 002
 Carretti E., O’Sullivan S. P., Vacca V., Vazza F., Gheller C., Vernstrom T., Bonafede A., 2023, *MNRAS*, 518, 2273
 Castorina E., Carbone C., Bel J., Sefusatti E., Dolag K., 2015, *J. Cosmol. Astropart. Phys.*, 2015, 043

- Cautun M., van de Weygaert R., Jones B. J. T., Frenk C. S., 2014, *MNRAS*, 441, 2923
 Cisewski J., Croft R. A. C., Freeman P. E., Genovese C. R., Khandai N., Ozbek M., Wasserman L., 2014, *MNRAS*, 440, 2599
 Cisewski-Kehe J., Fasy B. T., Hellwing W., Lovell M. R., Drozda P., Wu M., 2022, *Phys. Rev. D*, 106 023521
 Cole A., Shiu G., 2018, *J. Cosmol. Astropart. Phys.*, 2018, 025
 Cui W., Borgani S., Dolag K., Murante G., Tornatore L., 2012, *MNRAS*, 423, 2279
 de Graaff A., Cai Y.-C., Heymans C., Peacock J. A., 2019, *A&A*, 624, A48
 de Regt R., Apunevych S., von Ferber C., Holovatch Y., Novosyadlyj B., 2018, *MNRAS*, 477, 4738
 Dolag K., Bartelmann M., Lesch H., 1999, *A&A*, 348, 351
 Edelsbrunner H., Harer J. L., 2010, *Computational Topology*. American Mathematical Society, Providence, RI
 Elbers W., van de Weygaert R., 2019, *MNRAS*, 486, 1523
 Elbers W., van de Weygaert R., 2022, *MNRAS*, 520, 2709
 Euclid Collaboration, 2021, *A&A*, 647, A117
 Feldbrugge J., van Engelen M., van de Weygaert R., Pranav P., Vegter G., 2019, *J. Cosmol. Astropart. Phys.*, 2019, 052
 Figalli A., Glaudo F., 2021, *An Invitation to Optimal Transport, Wasserstein Distances, and Gradient Flows*. EMS Press, Berlin, Germany
 Frosini P., 1990, *Bull. Aust. Math. Soc.*, 42, 407
 Guzzo L. et al., 2014, *A&A*, 566, A108
 Heydenreich S., Brück B., Harnois-Déraps J., 2021, *A&A*, 648, A74
 Heydenreich S., Brück B., Burger P., Harnois-Déraps J., Unruh S., Castro T., Dolag K., Martinet N., 2022, *A&A*, 667, A125
 Huchra J. P. et al., 2012, *ApJS*, 199, 26
 Kahniashvili T., Maravin Y., Natarajan A., Battaglia N., Tevzadze A. G., 2013, *ApJ*, 770, 47
 Kimura Y., Imai K., 2017, *Adv. Space Res.*, 60, 722
 Libeskind N. I. et al., 2018, *MNRAS*, 473, 1195
 Lilly S. J. et al., 2007, *ApJS*, 172, 70
 Makarenko I., Shukurov A., Henderson R., Rodrigues L. F. S., Bushby P., Fletcher A., 2018, *MNRAS*, 475, 1843
 Maria C., Boissonnat J.-D., Glisse M., Yvinec M., 2014, in *Mathematical Software – ICMS 2014*. Springer, Berlin, Heidelberg, p. 167
 Pasquet J., Bertin E., Treyer M., Arnouts S., Fouchez D., 2019, *A&A*, 621, A26
 Planck Collaboration XIII, 2016, *A&A*, 594, A13
 Pranav P., 2021, preprint ([arXiv:2109.08721](https://arxiv.org/abs/2109.08721))
 Scodreggio M. et al., 2018, *A&A*, 609, A84
 Shao M., Anbajagane D., Chang C., 2022, preprint ([arXiv:2212.05964](https://arxiv.org/abs/2212.05964))
 Sylos Labini F., 2011, *Class. Quantum Gravity*, 28, 164003
 Sylos Labini F., Vasilyev N. L., Baryshev Y. V., 2007, *A&A*, 465, 23
 Tanimura H., Aghanim N., Kolodzig A., Douspis M., Malavasi N., 2020, *A&A*, 643, L2
 Tegmark M. et al., 2004, *ApJ*, 606, 702
 Tsizh M., Novosyadlyj B., Holovatch Y., Libeskind N. I., 2020, *MNRAS*, 495, 1311
 van de Weygaert R. et al., 2011, in Gavrilova M. L., Kenneth Tan C. J., Abolfazl Mostavi Mir, eds, *Astrophysics and Space Science Library*, Vol.6970, *Transactions on Computational Science XIV: Special Issue on Voronoi Diagrams and Delaunay Triangulation*. Springer, Berlin, Heidelberg, p. 60
 Vazza F., Feletti A., 2020, *Frontiers Phys.*, 8, 491
 Vazza F. et al., 2021, *Galaxies*, 9, 109
 Vernstrom T., Heald G., Vazza F., Galvin T. J., West J. L., Locatelli N., Fornengo N., Pinetti E., 2021, *MNRAS*, 505, 4178
 Wilding G., Nevenzeel K., van de Weygaert R., Vegter G., Pranav P., Jones B. J. T., Efstathiou K., Feldbrugge J., 2021, *MNRAS*, 507, 2968
 Xu X., Cisewski-Kehe J., Green S., Nagai D., 2019, *Astron. Comput.*, 27, 34

This paper has been typeset from a \LaTeX file prepared by the author.