Topological bias: how haloes trace structural patterns in the cosmic web

Raul Bermejo,[1]★ Georg Wilding [1,2,3]★ Rien van de Weygaert,[1,3]★ Bernard J. T. Jones [1] Gert Vegter[2,3] and Konstantinos Efstathiou[4]

[1]*Kapteyn Astronomical Institute, University of Groningen, PO Box 800, NL-9700 AV Groningen, The Netherlands*
[2]*Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University of Groningen, PO Box 800, NL-9700 AV Groningen, The Netherlands*
[3]*Centre for Data Science and Systems Complexity, University of Groningen, PO Box 800, NL-9700 AV Groningen, The Netherlands*
[4]*Division of Natural and Applied Sciences and Zu Chongzhi Center for Mathematics and Computational Science, Duke Kunshan University, No. 8 Duke Avenue, Kunshan 215316, Jiangsu Province, China*

ABSTRACT

We trace the connectivity of the cosmic web as defined by haloes in the Planck-Millennium simulation using a persistence and Betti curve analysis. We normalize clustering up to the second-order correlation function and use our systematic topological analysis to correlate local information and properties of haloes with their multiscale geometrical environment of the cosmic web (elongated filamentary bridges and sheetlike walls). We capture the multiscale topology traced by the halo distribution through filtrations of the corresponding Delaunay tessellation. The resulting nested *alpha shapes* are sensitive to the local density, perfectly outline the local geometry, and contain the complete information on the multiscale topology. We find a remarkable linear relationship between halo masses and topology: haloes of different mass trace environments with different topological signature. This is a *topological bias*, an environmental structure bias independent of the halo clustering bias associated with the two-point correlation function. This mass-dependent linear scaling relation allows us to take clustering into account and determine the overall connectivity from a limited sample of galaxies. The presence of topological bias has major implications for the study of voids and filaments in the observed distribution of galaxies. The (infra)structure and shape of these key cosmic web components will strongly depend on the underlying galaxy sample. Their use as cosmological probes, with their properties influenced by cosmological parameters, will have to account for the subtleties of topological bias. This is of particular relevance with the large upcoming galaxy surveys such as DESI, Euclid, and the Vera Rubin telescope surveys.

Key words: methods: data analysis – large-scale structure of Universe.

1 INTRODUCTION

Considering the Universe at a Megaparsec scale, the distribution of matter becomes anisotropic and inhomogeneous, and a highly geometric and connected structure emerges. This structure is known as the *cosmic web* (Bond, Kofman & Pogosyan 1996; Van de Weygaert & Bond 2008; Cautun et al. 2014). The cosmic web is, first and foremost, defined through its dark matter distribution. This distribution of densities is easy to obtain if one works on simulations, and allows a straightforward analysis of many of its features, as well as, for example, an identification of its constituent components. Observationally, the cosmic web is studied mostly through the spatial distribution of galaxies. It is, however, not clear in how far the structure traced by these galaxies is representative for the complete structure of the cosmic web.

We address this issue here: in this study, we focus on the cosmic web as traced by dark matter haloes, where we come upon the same problem – it is unclear how exactly the spatial pattern of dark haloes relates to the underlying matter distribution. As the spatial pattern of the cosmic web is determined by its constituents, which in turn are defined on the basis of their shape and mutual spatial connections, this naturally leads us to the field of topology. To shed light on the multiscale nature of the topology of the cosmic web traced by the population of dark haloes, we invoke persistent homology to decode the intricate features, patterns, and connectivity of the cosmic web.

Irrespective of the defining components, four notable qualities characterize the cosmic web. These are its anisotropy, the asymmetry between overdense and underdense regions, the connectivity between different morphological features, and its multiscale nature. The anisotropy of this web-like structure expresses itself through a rich morphology dominated by elongated filaments and tenuous walls. The asymmetry concerns the contrast between the overdense regions – clusters, filaments and a fair fraction of walls – and underdense regions. Together with cosmic walls and filaments, we see structure and morphological features over a wide range of scales, a direct manifestation of the bottom-up hierarchical build-up of the cosmic web and its structural components. Detailed theoretical discussions on the hierarchical evolution of the cosmic web have elucidated the intricate multiscale nature of the processes involved (see e.g. Bond & Myers 1996; Hanami 2001; Sheth & Tormen 2002; Sheth & van

de Weygaert 2004; Shen et al. 2006; Van de Weygaert & Bond 2008; Cadiou et al. 2020).

1.1 The connectivity and topology of the cosmic web

Of particular importance in the context of the topology of the cosmic matter distribution is the web-like network into which these morphological elements are organized. The resulting distinct connectivity finds its expression in clusters of galaxies at the junction of filaments, filaments at the intersection and boundaries of walls, and filaments and walls defining the boundaries of the large near-empty voids (Shandarin & Zeldovich 1989; Bond, Kofman & Pogosyan 1996; Van de Weygaert & Bond 2008; Aragón-Calvo, van de Weygaert & Jones 2010a; Aragón-Calvo et al. 2010b; Sousbie 2011; Sousbie, Pichon & Kawahara 2011; Cautun et al. 2014; Libeskind et al. 2018; Codis, Pogosyan & Pichon 2018a). In recent studies, the connectivity of the cosmic web in terms of the number of filaments that connect to a node has received considerable interest. This aspect of connectivity was first investigated by Aragón-Calvo et al. (2010b) on the basis of the MMF formalism (Aragón-Calvo et al. 2007a). The resulting dependence on the mass of the cluster nodes has recently been confirmed in a more profound and thorough analysis (Codis, Pogosyan & Pichon 2018a).

The connectivity is a key factor in the topological analysis of this study. To describe and characterize the multiscale connectivity of the cosmic web we turn to the mathematical characterization of connectivity: topology and persistent homology (Munkres 1984; Edelsbrunner & Harer 2010). In the study preceding this one, Wilding et al. (2021), we established the intimate link between the geometric structure of the cosmic web and its morphological elements and the topological signatures of the cosmic mass distribution. The surprising finding was that not only the structure of the cosmic web finds its expression in the topological parameters – Betti numbers and persistence diagrams – of the mass distribution but that even the key topological transitions such as the establishment of a connecting filamentary network are closely related to the dynamics of structure formation. Most telling is that the filaments start to define a percolating network at the density level corresponding to the gravitational binding of mass concentrations. On the basis of this intimate relationship between structure and topology, in the remainder of this study we take this as an established and known fact.

This study addresses the connectivity of the cosmic web as sampled by dark haloes. More specifically, we use Betti numbers (Poincaré 1892) and persistence diagrams to study the connectivity of the spatial distribution of dark haloes. In this study, we do so by assessing the spatial distribution of dark matter haloes in a state-of-the-art dark matter simulation marked by a large cosmic volume and vast dynamic range. The Planck Millennium [P-Millennium hereafter, see Ganeshaiah Veena et al. (2018) and Baugh et al. (2019)] contains 7.7×10^7 of these haloes.

Here we focus on the topology, and in particular connectivity, of the cosmic mass distribution as traced by the discrete population of haloes. The dark matter haloes, and implicitly the galaxies that reside in their interior, populate the components of the cosmic web. It means that the spatial pattern outlined by dark matter haloes constitutes a diluted and possibly biased reflection of the underlying web-like dark matter distribution. We investigate the dependence of the web-like distribution of dark haloes as a function of their intrinsic properties, in particular that of their mass. This involves an analysis of the patterns and topological features and characteristics – Betti numbers and persistence – of the web-like structure delineated by dark matter haloes. We assess these properties for the spatial distribution of haloes

Table 1. Halo subpopulations in the P-Millennium simulation. For each halo class, we include the logarithmic mass range and the approximate number of haloes for each population.

Halo population	$\log_{10}(M)$ $[h^{-1}\mathrm{M}_\odot]$	N (number of haloes)
HA	(10.5,11]	$\simeq 7.05 \times 10^6$
HB	(11,11.5]	$\simeq 2.67 \times 10^6$
HC	(11.5-12]	$\simeq 9.69 \times 10^5$
HD	(12,12.5]	$\simeq 3.47 \times 10^5$
HE	(12.5,13]	$\simeq 1.21 \times 10^5$
HF	(13,13.5]	$\simeq 4.05 \times 10^4$

in six mass ranges. The aim is to decode the topological embedding of the haloes, i.e. the relation between the morphological components of the cosmic web traced by haloes, their topological characteristics and connectivity, and the mass of the haloes.

P-Millennium provides a very useful halo catalogue for the topological analysis presented in this study because of its broad mass range of haloes. Haloes vary in masses of $\log_{10}(M) = [10.5, 13]\,h^{-1}\mathrm{M}_\odot$, allowing to resolve any topological differences between haloes more precisely. To complete such a task, we divided the halo catalogue into six mass bins. Throughout the rest of the paper, these bins are labelled as HA, HB, HC, HD, HE, and HF. Their associated mass ranges and number of haloes are shown in Table 1.

As a basis to analyse the topology of these haloes, we use a distance-based filtration. In this kind of filtration, haloes and clusters of haloes form connections if closer than a specific distance. This way of defining connections on a point-based distribution (the dark matter haloes) is formalized through a Delaunay tessellation (Delaunay 1934) and the resulting simplicial complex, which represents the halo structure as a collection of simplices. The simplices are naturally adaptive to the local density and geometry. Choosing a specific filtration length yields a uniquely defined scale-dependent subset of the full Delaunay complex, the so-called *alpha shapes*, introduced by Edelsbrunner and collaborators (Edelsbrunner, Kirkpatrick & Seidel 1983; Edelsbrunner & Mücke 1994; Edelsbrunner & Harer 2010). The *alpha complex*, the nested sequence of alpha shapes, perfectly traces the multiscale connections established by the halo distribution, while considering that the resulting connections occur on all scales. The sequence forms a direct and unambiguous representation of the multiscale topological structure of the distance field. The derived persistence diagrams and Betti numbers characterize the underlying homology of the alpha shapes of the halo distribution in this simulation. Through the use of persistence computed from the alpha shape filtration, we establish a natural description of the multiscale nature of the halo distribution.

The applications of persistent homology (Edelsbrunner & Harer 2010), Betti numbers (Poincaré 1892), and topological data analysis in general have seen a large increase in number and popularity in recent years (for a review see Wasserman 2018). We see the popularity also in a diverse range of application fields, ranging from brain research (Petri et al. 2014; Reimann et al. 2017) to materials science (Hiraoka et al. 2016), and in recent years also to cosmology and astrophysics. Over the past decade, an increasing number of cosmological studies have invoked persistent homology in an attempt towards quantifying and classifying the complex web-like patterns in the cosmic matter and galaxy distribution (Van de Weygaert et al. 2010; Sousbie 2011; Sousbie, Pichon & Kawahara 2011; Van de Weygaert et al. 2011; Shivashankar et al. 2016; Kimura & Imai 2017; Pranav et al. 2017; Feldbrugge et al. 2019; Kono et al. 2020; Biagetti, Cole & Shiu 2021). Additional interesting applications concern the

investigation of the evolving topology of the reionization network (Elbers & van de Weygaert 2019; Giri & Mellema 2021; Thélie et al. 2021) and the exploration of the structure and topology of magnetic fields in the interstellar medium (Makarenko et al. 2018).

1.2 Environmental influences on the formation and evolution of haloes and galaxies

The dependence of the properties of haloes of galaxies, and of their formation and evolution, on the large-scale environment is a vigorous issue of study and discussion in the cosmological community. The first direct indication for a major systematic influence is the discovery by Dressler (1980) of the density-morphology relation. Early-type galaxies are predominantly found in high density regions, in particular clusters of galaxies, while late-type galaxies form the majority of galaxies in more moderate density environments. A particularly nice illustration of this in the context of the web-like galaxy distribution is the segregation between ellipticals and spirals in the Pisces-Perseus supercluster (Giovanelli, Haynes & Chincarini 1986).

There are indications for a significant impact of the large-scale environment on a range of properties of galaxies and haloes. Particularly outstanding is the relation between galaxy properties and that of density of the environment, of which the morphology-density relation is the most well-known representative (Dressler 1980). The influence of the location in the various morphological environments in the cosmic web – clusters, filaments, walls and voids – on halo and galaxy properties appears to be more subtle. Also, it remains an as yet unsettled issue whether the local density is the sole dominant factor, or whether the morphological nature of the environment also represents a significant supplementary or modulating influence (Goh et al. 2019; Hellwing et al. 2021). However, to some extent this concerns an ill-defined dichotomy: the density of the environment is augmented by the anisotropy of its gravitational contraction and collapse (Icke 1973; Bertschinger & Jain 1994; Bond & Myers 1996).

Many studies indicate that the paramount factor determining the physical nature of a galaxy and its dark matter halo is the mass of the halo. The mass of haloes is itself intimately linked to their environment in the cosmic web: the halo and galaxy mass functions are strongly determined by whether the haloes reside in filaments, walls, voids or cluster nodes (Cautun et al. 2014; Ganeshaiah Veena et al. 2018, 2019). While the number density of haloes in filaments, harbouring 50 per cent of dark matter, haloes and galaxies, is somewhat higher than the average density, it is much lower in walls and voids. The mass function in walls and voids is also shifted considerably to lower masses: walls, and even more so voids, are populated by much smaller haloes and galaxies than those populating the filaments, while cluster nodes contain the most massive haloes and galaxies. This is also reflected in the dependence of halo clustering as well as halo bias on halo mass (Yang et al. 2017), and/or their proximity to morphological features of the cosmic web (Kraljic et al. 2018).[1]

Arguably, the most outstanding influence of the cosmic web on the formation and evolution of galaxies is that of rotation. The tidal force field induced by the evolving inhomogeneous mass distribution is responsible for the anisotropic collapse of filaments and walls, as well as for the torquing of contracting mass clumps (Hoyle, Burgers & van De Hulst 1949; Peebles 1969; White 1984; Porciani, Dekel &

Hoffman 2002b; Schäfer 2009). This induces a significant correlation between the direction of filamentary ridges and the spinning direction of collapsed haloes (Lee & Pen 2001; Porciani, Dekel & Hoffman 2002b). The alignment between filaments and galaxy spin has also been found in observations (Jones, van de Weygaert & Aragón-Calvo 2010; Tempel & Libeskind 2013; Van de Sande et al. 2021). Starting with the work by Aragón Calvo (2007), a large number of studies have shown that the halo spin alignment with respect to large-scale filaments is mass-dependent (Aragón Calvo 2007; Hahn et al. 2007; Zhang et al. 2009; Codis et al. 2012; Trowland, Lewis & Bland-Hawthorn 2013; Wang & Kang 2017; Ganeshaiah Veena et al. 2018; Codis et al. 2018b), which may be an indication for the influence of anisotropic inflow along the filamentary arteries (Ganeshaiah Veena et al. 2018, 2019, 2021; López et al. 2021). The reality of this effect has recently been confirmed by observations (Tempel & Libeskind 2013; Welker et al. 2020).

A major additional influence of the cosmic web on the properties of haloes and galaxies is that on their growth and assembly history. A range of studies have indicated that the growth of haloes is significantly influenced by the cosmic web environment. The relation between halo growth and cosmic web environment is partially a reflection of the tidally directed dynamical evolution of and anisotropic inflow on to the haloes (Aragón-Calvo et al. 2007b; Dalal et al. 2008; Hahn et al. 2009; Borzyszkowski et al. 2017; Musso et al. 2018; Paranjape, Hahn & Sheth 2018; Zhang, Yang & Guo 2021). The result is a halo assembly history and time that is modulated by the cosmic web, and translates into a halo and galaxy bias known as assembly bias (Gao, Springel & White 2005; Wechsler et al. 2006; Dalal et al. 2008; Mao, Zentner & Wechsler 2018). Salcedo et al. (2020) found observational evidence for such bias in the SDSS survey, while it may also be reflected in the relation between the brightness of galaxies and their level of clustering as has been established for the GAMA survey (Jarrett et al. 2017) and SDSS survey (Paranjape, Hahn & Sheth 2018).

1.3 Topological bias: identification and quantification

The central question of this study is whether the topological patterns and connections we detect through persistent homology *also* involve a dependence on the intrinsic properties of the tracing haloes. If so, this implies a dependence that we define as *topological bias*.

Consider the subsample defined by the more massive haloes: it is relatively sparse and has lower density than samples of lower mass haloes. This situation is directly reflected in the larger scale length of the two-point correlation function for the different mass class haloes. The different mass samples have greatly different density. The detection of a mass-dependent topological bias requires that we compensate for this. To achieve this compensation, the spatial scales of the topological features in a given mass class are rescaled using the correlation length of that class. After this renormalization the connected structures of halo populations of different mass ranges (and accordingly vastly different number densities) can be compared and investigated on scales of similar length and *equal clustering*.

Accordingly, before running the topological analysis we renormalize the scale of each sample relative to its clustering length-scale r_0. The details of this analysis are presented in Section 4.2.

Deviations in topological properties of the renormalized halo distribution are implicit indications and manifestations of the presence of higher order clustering. They would elude detection through a two-point correlation function analysis, and are a direct indication for complex topological influences.

---

[1]The definition of proximity to a filament in the cosmic web is to some extent predicated by the particular classification technique used. A method that allows the identification of filaments of varying thickness would relate galaxy properties to the thickness and prominence of the filament.

We observe that haloes in different mass ranges trace specific structural patterns, characterized by topological features of varying dimension, at different *normalized* length-scales. This is a direct indication and compelling evidence for the presence of a bias that involves nontrivial contributions by higher order clustering terms. This topological bias is found to have a simple dependence on the mass of the dark matter haloes that define the structure. Moreover, and perhaps surprisingly, we find a clear systematic behaviour of the topological bias: there is a strict relation between the normalized topological parameters and the characteristics of the halo population.

1.4 Implications of topological bias

Topological bias addresses the question in how far the morphology of voids and filaments, and the corresponding tunnels, depend on the nature of the halo populations. It means that the cosmic web traced by haloes in different mass ranges exhibits differences in the corresponding topological characteristics, such as the connectivity properties. More generally, it also involves aspects such as volume, shape and substructure of these features. In other words, it implies a systematic dependence of the spatial structure of the cosmic web on the specifics of the halo population that is used as its tracer. Moreover, when normalized and compensated for two-point clustering, topological analysis reveals that the structure of the cosmic web traced by the different halo populations displays fundamental differences. In other words, filaments and voids for different halo populations do have significant differences. These go further than simple self-similarity and includes effects of substructure, connectivity and shape.

The innately multiscale nature of persistent topology provides new insights into the hierarchical nature of the structure formation process. Structural features are the product of a process marked by a rich assembly history. This tends to leave traces in the (sub)structure of such features, to which the multiscale topological analysis in terms of homology and persistence is very sensitive. We therefore expect such an analysis to highlight the link between the assembly history and cosmic environment. This may imply an intricate and close relationship between the topological bias which we find and assembly bias (Gao, Springel & White 2005; Wechsler et al. 2006; Mao, Zentner & Wechsler 2018).

Observing topological bias will have a substantial impact on the information on the cosmic web that can be gained from cosmological studies. The detection and incorporation of topological bias in the analysis of cosmological data sets is crucial if one seeks to exploit higher levels of structural complexity in the distribution of galaxies and galaxy haloes. This concerns important aspects such as the properties of the void population, that of the filamentary spine of the cosmic web and its branching network of tendrils. The visually more directly accessible description in terms of topological factors, such as Betti numbers and persistence, also provides a considerably better appreciation of the nature of the contributing higher order clustering terms. Topological bias may also imply complications when seeking to use the observed galaxy distribution to extract information on the underlying dark matter distribution and on general cosmological parameters. The presence of topological bias reveals systematic differences between the structural pattern outlined by the dark matter distribution and that by galaxies.

1.5 Programme and outline

To unravel the relationship between the topology of the galaxy distribution and the underlying dark matter distribution, we first

have to establish a systematic, preferably quantifiable, relationship between the topological characteristics of the spatial patterns outlined by the various populations. This study limits itself to an investigation of topological properties of populations of dark matter haloes in the P-Millennium simulation, in a subsequent study we will extend this to that of the simulated galaxy distribution in the IllustrisTNG simulations. In a final stage, we will extend our analysis to the observed galaxy distribution. A range of available redshift surveys, e.g. the 2dFGRS, SDSS, GAMA, 2MASS, and VIPERS (Colless et al. 2003; Tegmark et al. 2004; Driver et al. 2009; Huchra et al. 2012; Guzzo & VIPERS Team 2013), already produced detailed three-dimensional maps that revealed the presence of a rich web-like morphology and topology. The very recent Year 3 data release of DES (Sevilla-Noarbe et al. 2021) and the upcoming EUCLID survey (Laureijs et al. 2011) as well as the DESI experiment (DESI Collaboration 2016; Dey et al. 2019) will provide ample opportunity for the identification of topological features to analyse their multiscale (sub)structure, and for applications of the toolset of topology and persistent homology described in this study.

In this paper, we extend the earlier work of the group laid out in Van de Weygaert et al. (2011), Nevenzeel (2013), Pranav et al. (2017, 2019a,b), Feldbrugge et al. (2019), and Wilding et al. (2021). It represents a key step in enabling the formalism developed in these previous studies towards one suited for the analysis of galaxy redshift surveys. To this end, this paper is structured as follows: in Section 2, we introduce the formalism of persistent homology, which is necessary to understand the construction of alpha shapes that yield the persistence diagrams and Betti curves. In Section 3, we describe how we obtained the halo catalogue that was used in the study. In this section, we also motivate the mass binning we applied to the catalogue and describe some of its halo properties, such as spatial distribution and clustering. Section 4 presents the results of this study, mainly the intensity persistence diagrams and Betti curves which aim to support our claim of a topological bias. Moreover, in this section we discuss the implications of our intensity persistence diagrams and Betti curves, and explain how these results provide compelling evidence for the existence of a topological bias. Finally, in Section 5, we summarize the results of this paper and indicate some future research prospects following this work.

2 TOPOLOGY, PERSISTENCE, AND ALPHA SHAPES

Topology is the field of mathematics that studies properties which are preserved under continuous deformations (such as bending and stretching, but not gluing). Connectivity is a key *topological invariant* (Sutherland 2009) which studies mathematical features in a given point distribution. Formally, these topological features are structures that prevent the space to be continuously deformed into a single point. The canonical probe of connectivity is that of counting topological features, and establish their relationship with neighbouring features.

The most outstanding virtue of our analysis is that it allows us to evaluate the prominence of the *global* topology of the different topological features as a function of spatial scale, density or other relevant factors. This is of major importance for our understanding of the structure of the cosmic web, and of its hierarchical assembly.

We are also interested in a direct assessment of the dependence of the structure of the mass distribution on spatial scale as inferred from the observed sample of objects. The multiscale nature of the matter distribution would follow from the assessment of changes in the field structure as a function of distance (level).

2.1 Topological features

In the context of this study, we distinguish three different topological features. These are associated with (super)clusters, filaments and voids. Each of these features correspond to a well-defined topological class with corresponding dimensionality. Throughout the rest of the paper, we will be referring to (super)clusters, filaments – and their associated tunnels – and voids as *topological features*.

Formally, the dimensionality is associated with the type of topological hole that each feature corresponds to. In the context of this study, zero-dimensional holes correspond to the space separating two disconnected (super)clusters, a one-dimensional hole is associated with a closed loop of filaments, and a two-dimensional hole corresponds to a void bounded by a closed surface. It is the study of these topological holes and their boundary that provides a formal notion of connectivity in the distance field that we study. For example, zero-dimensional holes can also refer to the space separating two *(super)clusters* of haloes.

2.2 Homology

The formalism of homology allows us to probe the number of zero-, one- and two-dimensional structural features and their connectivity in the underlying matter distribution. The number of the various topological features is expressed in terms of *Betti numbers* (Poincaré 1892). In proper mathematical formulation, the *d*th Betti number β_d corresponds to the rank of the *d*th homology group. An in-depth discussion of the theory of homology groups lies outside the scope of this paper. For the interested reader, we refer to Vegter (2004); Zomorodian & Carlsson (2005); Rote & Vegter (2006); Edelsbrunner & Harer (2010) and Robins (2006); Robins (2015) for a general introduction of homology theory and its application in computational topology. For the more specific case of Betti numbers and persistent homology in the context of the density field of the cosmic web, we refer to, for example, Van de Weygaert et al. (2010, 2011), Sousbie (2011), Pranav et al. (2017, 2019a), Xu et al. (2019), and Wilding et al. (2021).

2.3 Betti numbers and cosmic web topology

Intuitively, the Betti numbers count the number of topological features in the space. That is, in the context of haloes outlining the cosmic web, β_0 counts the number of disconnected clusters of haloes, β_1 counts the number of filamentary loops enclosing independent tunnels and β_2 counts the number of shells enclosing isolated voids. More precisely, the *d*th Betti number corresponds to the number of *d*-dimensional holes of the sampled space.

Many studies have probed the topology of the cosmic mass and galaxy distribution with the *Euler characteristic χ* (Euler 1758; Adler 1981; Gott, Melott & Dickinson 1986; Hamilton, Gott & Weinberg 1986; Park et al. 2013; Pranav et al. 2019a), often via the directly related *genus*. The Euler characteristic represents a profound topological invariant, which finds itself at the junction of several branches of mathematics, including homology and simplicial topology (see Adler 1981; Adler & Taylor 2010; Pranav et al. 2019a). The Euler characteristic of a three-dimensional set is the number of its connected components, minus the number of its tunnels, plus the number of voids it contains. It is also a geometric quantity via the Gauss-Bonnet theorem, the deep and seminal expression going back to Euler (Euler 1758). Requiring both Differential and Algebraic Topology, it shows that the Euler characteristic also has a geometric interpretation, and is associated with the integrated Gaussian curvature of a manifold. In fact, together with other quantities related to volume, area and length, the Euler characteristic forms a part of a more extensive geometrical descriptions in terms of *Minkowski functionals* or Lipschitz-Killing curvatures. These have been invoked in a cosmological context in a range of studies (see e.g Mecke, Buchert & Wagner 1994; Schmalzing & Buchert 1997; Sahni, Sathyaprakash & Shandarin 1998; Schmalzing & Gorski 1998; Schmalzing et al. 1999; Kerscher, Szapudi & Szalay 2000).

There is a profound relationship between the homology characterization in terms of Betti numbers and the Euler characteristic (see Pranav et al. 2019a). The *Euler-Poincaré* formula states that the Euler characteristic is the alternating sum of the Betti numbers (also see e.g. Van de Weygaert et al. 2011; Pranav et al. 2017, 2019a),

$$\chi = \beta_0 - \beta_1 + \beta_2 . \tag{1}$$

This expression immediately reveals that the Betti number characterization of the web-like halo distributions represents a more elaborate and visually insightful description of the topology of the cosmic web than that quantified only in terms of the Euler characteristic or genus. Visually imagining the 3D situation as the projection of three Betti numbers on to a one-dimensional line, we may directly appreciate that two manifolds that are branded as topologically equivalent in terms of their Euler characteristic may actually turn out to possess intrinsically different topologies when described in the richer language of homology. Evidently, in a cosmological content this will lead to a significant increase of the ability of topological analyses to discriminate between different cosmic structure formation scenarios.

2.4 Multiscale topology and filtrations

While the Betti number analysis provides a global inventory of structural and topological changes as a function of scale, homology entails a considerably richer source of information with respect to the multiscale character of the mass distribution. It allows the identification and characterization of individual topological transitions. As such, it provides a direct means of studying the means by which individual features are establishing connections with neighbouring features. The explicit definition in terms of the formalism of *Persistent Topology* (Edelsbrunner & Harer 2010) provides us with a highly informative means of quantifying the intricate connectivity of the cosmic web, and hence represents the desired characterization in terms of a solid mathematical foundation.

The key element of this study concerns the change of the number of topological features as a function of spatial scale. As the scale changes, new features may emerge. This may involve the merging of formerly disconnected features, as well as the disappearance and annihilation of features. Examples of the latter happen when tunnels or cavities fill up. Such topological changes happen whenever the change in scale leads to the inclusion or removal of critical points, i.e. the minima, maxima or saddle points of the underlying manifold.

2.4.1 Filtrations

To facilitate persistence analysis, for the analysis of the multiscale nature of the mass distribution, the first step is a systematic definition of structural scale dependence. This leads to the concept of *filtration*. It involves the quantification of structure in terms of a field *S* that entails the characteristics of the mass distribution.

To establish the topological connections of the mass distribution, it is important to study the changes in topological structure as we proceed along different levels of the field *S*. This involves the emergence, merging, and disappearance of individual topological features, indicated as the *birth* and *death* of these features.

Mathematically, the set of filtrations of the field S is the nested sequence of subsets S_j for varying field levels:

$$\emptyset = S_0 \subseteq S_1 \subseteq S_2 \subseteq ... \subseteq S_M = S. \tag{2}$$

Although the choice of filtration scale can allow for continuous values, a key observation is that we are only dealing with a finite number of critical points, and thus effectively with a finite number of filtrations.

In a cosmological context, relevant filtrations involve a sequence of density field levels or a spectrum of spatial scales. In the case of the latter, for a discrete point sample we may accomplish this in terms of the (continuous) distance field.[2] Another option would be to circumvent the need for the calculation of the distance field, and seek to work directly from the object sample. It is the latter option which we pursue in this study.

2.4.2 Density field filtrations

In earlier works of our group (Pranav et al. 2017, 2019a,b; Feldbrugge et al. 2019; Wilding et al. 2021), we assessed the multiscale topology of the cosmic mass distribution by means of the *density filtrations*, where connections are formed based on the local density values. In this case, the field filtration is defined by the nested sequence of superlevel sets of the density field $f(\mathbf{r})$ by applying a threshold ν to the function values. The corresponding superlevel set $M(\nu)$ is

$$M(\nu) = \{(\mathbf{r}, f(\mathbf{r})) \mid f(\mathbf{r}) \geq \nu\}, \tag{3}$$

with the key *nestedness property*

$$M(\nu_1) \subseteq M(\nu_2) \text{ for } \nu_1 \geq \nu_2. \tag{4}$$

In view of this nestedness property, the superlevel sets $M(\nu)$ form a *filtration*, i.e. a hierarchically ordered collection of nested sets as the density decreases from $+\infty$ to 0.

2.4.3 Distance field filtrations

Another natural choice for the filtration of a field sampled by a discrete number of sample points is that of the distance field. It is defined as the value of the distance to a sample point, usually the closest sample point (see e.g. Aragón-Calvo, van de Weygaert & Jones 2010a). The focus on the distance field enables a more straightforward identification of features on a certain spatial scale: by defining the filtration of the field with respect to the distance values and grouping the sample points that get connected at a given distance threshold, we effectively identify structural features at the corresponding scale. For more detailed information on simplicial complexes, we refer to Appendix A.

Hence, the nested sequence of the distance field filtration will result in a complete scale-space representation of the mass distribution (Aragón-Calvo et al. 2007b; Cautun, van de Weygaert & Jones 2013). It will allow a detailed assessment of how structures at different spatial scales relate to each other and connect up in the cosmic mass distribution.

Because the distance field is defined with respect to the discrete point distribution, we may actually not need the computation or reconstruction of a continuous distance field. Instead, we may work directly from the point distribution. To that end, the points are spatially connected by means of a *Simplicial Complex* that reflects the density, geometry and multiscale character of the spatial point distribution.

2.5 Simplicial topology: Delaunay triangulations and alpha shapes

A simplicial complex is an ordered geometric assembly of faces, edges, nodes, and cells that constitutes a discrete spatial map of the volume containing a point set (Edelsbrunner & Harer 2010; Pranav et al. 2017). The geometric components of the complex are simplices of different dimensions: a cell is a three-dimensional simplex, a face or wall is a two-dimensional simplex, an edge is a one-dimensional simplex, and a node is a zero-dimensional simplex.

A good impression of a typical two-dimensional simplicial complex can be obtained from Fig. 1, showing a slice through a Delaunay tessellation (bottom right-hand panel) and a series of *alpha shapes* (see Appendix B). Zero-, one- and two-dimensional simplices are clearly visible: they correspond to the vertices and edges of the particle distribution's tessellation.

The simplicial complexes at the centre of this study are Delaunay and Voronoi tessellations (Voronoi 1908; Delaunay 1934).

The self-adaptive nature of the Delaunay tessellations also assures us that it represents the topological structure traced by the sample points. Given the intention to assess the multiscale topological aspects of the mass distribution probed by the sample points, it leads us to consider the possibility to define a scale-based filtration of the Delaunay tessellation that is the simplicial representation of the probed mass distribution. The filtration of a simplicial complex K is given by

$$\emptyset = K_0 \subseteq K_1 \subseteq K_2 \subseteq ... \subseteq K_m = K. \tag{5}$$

The hierarchical subsets K_i of K encode the change in topology throughout the simulation box. It naturally leads to the concept of the scale-based filtration of a Delaunay tessellation, known as *alpha shapes* (Edelsbrunner, Kirkpatrick & Seidel 1983; Edelsbrunner & Mücke 1994; Edelsbrunner & Harer 2010). Uniquely defined for a particular point set P by a real number α, the scale parameter, alpha shapes correspond to a unique assembly of simplicial (geometric) structures that capture the shape and morphology of the point distribution. They are a well-known concept in Computational Geometry and Computational Topology, and involve a generalization of the convex hull of a point set. For detailed information on and illustrations of alpha shapes see Appendix B.

2.6 Persistent homology on alpha shapes

Instead of studying the topology at a single length-scale, alpha shape topology probes the topology of the distance field along its set of filtrations as a function of the length-scale α.[3] As the filtration length-scale α increases, the halo/sample point distribution proceeds through a process in which we see a growing number of Delaunay simplices getting connected as they get embedded in

---

[2]Note that there is a strong correlation between the distance field values and the density field: short distance field values correspond to regions of high density.

[3]Alpha shapes represent a mathematically well-defined filtration of the distance field, with a direct visual connection to the spatial pattern analysed. There are a range of other formalism and methods for the analysis of aspects of spatial point distributions that are based on the corresponding distance field. Amongst others, Minimal Spanning Trees and graph theoretical techniques are known for their application to the large-scale galaxy distribution. In Appendix C, we include a short discussion and comparison of several distance field based methods.
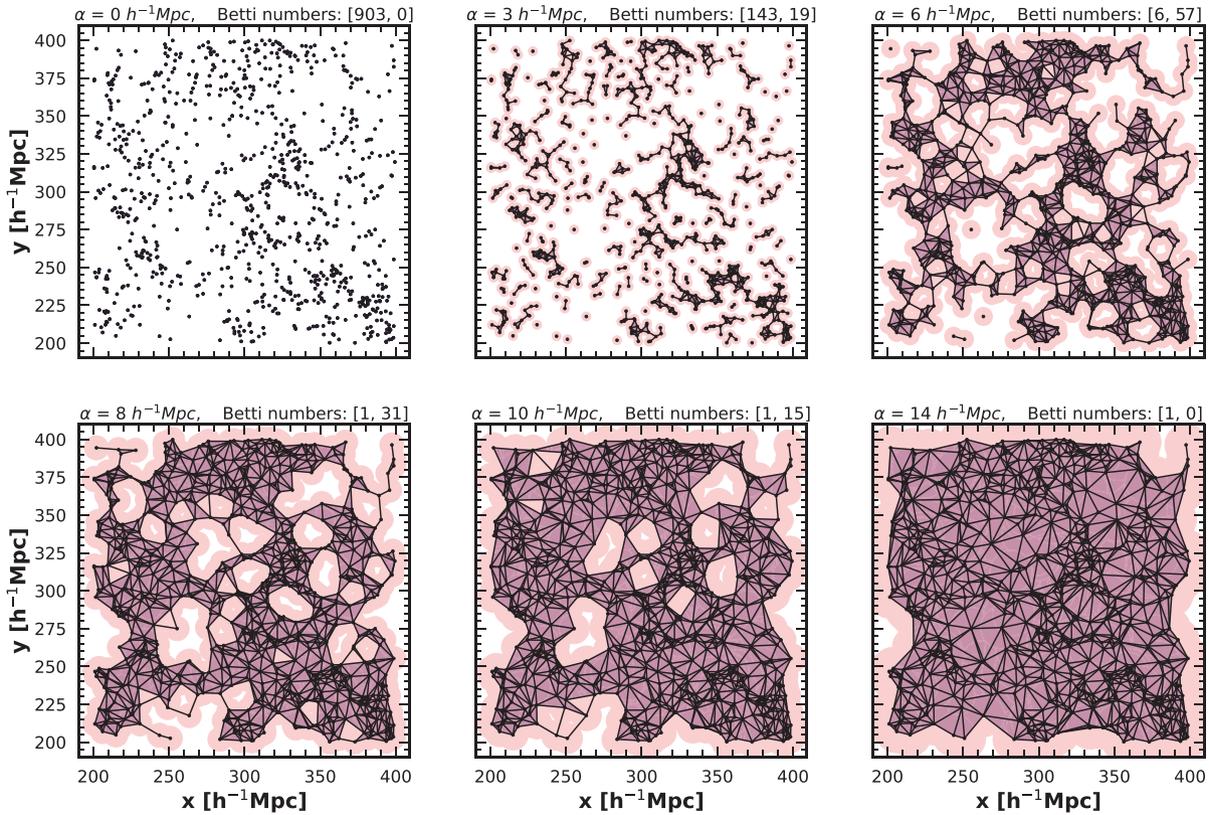
Figure 1. Alpha shape of a slice of haloes in the P-Millennium. This plot represents the evolution of a two-dimensional alpha shape for a sample of haloes with a mass range of $\log_{10}(M) = (12.5, 13]h^{-1}\mathrm{M}_\odot$ in a slice of $30h^{-1}\mathrm{Mpc}$ from the halo catalogue. Different panels are associated with different values of α, representing different stages in the evolution of the alpha shape. From left to right in each row: $\alpha = 0\,h^{-1}\mathrm{Mpc}$, $\alpha = 3\,h^{-1}\mathrm{Mpc}$, $\alpha = 6\,h^{-1}\mathrm{Mpc}$, $\alpha = 8\,h^{-1}\mathrm{Mpc}$, $\alpha = 10\,h^{-1}\mathrm{Mpc}$, $\alpha = 14\,h^{-1}\mathrm{Mpc}$. These plots show how, as we increase the value α, topological features (such as tunnels) are born and disappear and the haloes become more connected. The last panel represents an almost complete version of the Delaunay tessellation.

the corresponding alpha complex. As a result of this, we see a continuously changing population of different separate simplicial complexes and the corresponding structural entities they represent.

Proceeding through the entire value range of the scale parameter α, we may follow the creation and destruction of the individual topological features. As new structures emerge, existing structures merge and other structures get annihilated, the number of (super)clusters of haloes, filaments and voids will continuously change. To obtain insight into the connection between the growing simplicial (alpha shape) complex and the topology of the spatial structure, it is illustrative to consider what happens as more points get added while the simplicial complex grows (see Van de Weygaert et al. 2011). Cycling through the simplices of the alpha complex, we find the following:

When a point is added to the alpha complex, a new component is created: the zeroth Betti number β_0 is increased by 1. In this study, the alpha complex starts out with all haloes added at $\alpha = 0$, and each component corresponds to an individual halo. When adding edges between pairs of points, the components become (super)cluster of haloes. The process of adding edges between points can have two outcomes. Either both points belong to the same, or to different components of the current complex. When they belong to the same component, the edge creates a new tunnel, increasing the one-dimensional Betti number by 1. The creation of a filament corresponds to the creation of a tunnel. When the edge connects two different components, two (super)clusters merge and β_0 decreases by 1. When a triangle gets added, it may complete the enclosure of a void or it may close a tunnel. In the first case, a new void is created

and β_2 increases by 1. In the latter case, β_1 is decreased by 1 as the number of tunnels and corresponding filaments decreases. Finally, when a tetrahedron is added, a void is filled. This means that the two-dimensional Betti number β_2 is lowered by 1.

By assessing the value of α at which a feature is created, the *birth value* α_b, and the value at which a feature is destroyed, the *death value* α_d, we follow the existence of (super)cluster complexes, of filamentary features and of enclosed voids (see, e.g. Wilding et al. 2021). The difference between the creation and the destruction value,

$$\pi : = \alpha_d - \alpha_b \qquad (6)$$

denotes the *persistence* π of a topological feature. By plotting the values of α_d versus α_b for all topological features present in a spatial point distribution, for each dimension $d = 0, 1, \ldots, D$, a highly detailed and idiosyncratic depiction of the topological structure is obtained. These *persistence diagrams* not only quantify the topology in terms of a few summarizing parameters, but in terms of diagrams that contain information on every individual topological feature present in the sample point distribution.

Analogously, we can compute the Betti numbers (β_0, β_1 and β_2) of the distance field as a function of the length-scale α. The resulting Betti curves (i.e. the Betti numbers for all scales) provide an overview on the complete structure of critical points, and on how the global topology differs for varying scales. While Betti curves inform us of the overall topology (since they measure the total number of topological features as a function of α), persistence diagrams allow

us to precisely track at what length-scales each of these topological features is formed and for how long they *persist*.

2.7 Topological data analysis

In order to obtain the persistence points and Betti numbers from the formalism described above, we used GUDHI (Geometric Understanding in Higher Dimensions). GUDHI is a generic open source C++ library for Computational Topology and Topological Data Analysis (The GUDHI Project 2015).

We start from the halo distribution in the P-Millennium, sampled for a specific mass bin, and treating it as point-cloud data. GUDHI allows the straightforward computation of the alpha complex (Rouvreau 2015), making use of the corresponding CGAL [the Computational Geometry Algorithms Library; The CGAL Project (2019)] procedure. The simplices of the alpha complex are stored in a simplex tree (Maria 2015a), each with the corresponding filtration α. The persistent homology, i.e. the persistence points (α_i, α_j), is then calculated using a persistent cohomology algorithm (Maria 2015b). For a more detailed description of the algorithm, we refer to Boissonnat & Maria (2012) and The GUDHI Project (2015). The Betti numbers that we will use to analyse the halo structure can be calculated directly from the persistence points. This is done for a chosen α by counting the pairs that were *born* at a lower α and will *die* at a higher α.

3 HALO POPULATION

3.1 Simulation and halo finding

In order to study the connectivity of haloes in the cosmic web, we use the state of the art *N*-body simulation for a standard ΛCDM model: the Planck-Millennium simulation [hereafter P-Millennium, see Ganeshaiah Veena et al. (2018) and Baugh et al. (2019)]. The P-Millennium traces the cosmic evolution of over 128 billion (5040^3) dark matter particles of mass $M = 1.061 \times 10^8 \, h^{-1}\mathrm{M}_\odot$ in a periodic box of side length $542.16 \, h^{-1}\mathrm{Mpc}$. The P-Millennium compromises cosmological parameters from the latest Planck survey results (Planck Collaboration 2018): the Hubble-Lemaître parameter $h = 0.6777$, where $h = H_0/100 \, \mathrm{km \, s^{-1} \, Mpc^{-1}}$ and H_0 is the Hubble-Lemaître constant. Density parameters correspond to $\Omega_m = 0.307$, $\Omega_\Lambda = 0.693$, and the density fluctuation is $\sigma_8 = 0.8288$. The evolution of these dark matter particles is traced back to $z = 127$. However, in this work, we limit ourselves to the study of haloes in the P-Millennium at present time $z = 0$.

The halo catalogue was obtained by running the halo and subhalo finder SUBFIND (Springel et al. 2001). Subsequently, halo merger trees were obtained by using the DHALOS algorithm described in Jiang et al. (2014). A more detailed description of how the halo catalogue was obtained can be found in Ganeshaiah Veena et al. (2018) and Baugh et al. (2019). We define the halo radius R_{200} as the radius of a sphere enclosing a mass M_{200}. That is, each halo encloses an overdensity corresponding to $\Delta = 200$ times the critical density of the Universe. In total, there are approximately P-Millennium.

In this study, we consider haloes more massive than $3.2 \times 10^{10} \, h^{-1}\mathrm{M}_\odot$, corresponding to 1.13×10^7 of such haloes. This filtering has been chosen so that each halo is resolved by a large enough number of dark matter particles (300 particles, for more details see Bett et al. 2007; Ganeshaiah Veena et al. 2018). Nonetheless, our sample of 1.13×10^7 is large enough to characterize the topology of the cosmic web in a statistically representative manner. The haloes vary in mass from $\log_{10}(M) = [10.5, 13] \, h^{-1}\mathrm{M}_\odot$. The halo catalogue is into six mass bins, labelled as HA, HB, HC,

HD, HE, and HF. The mass ranges and number of haloes in these bins are shown in Table 1.

3.2 Mass dependence halo clustering

Fig. 2 shows the distribution of a subsample of haloes of different mass ranges in the P-Millennium. As we consider less massive haloes, some morphological elements of the cosmic web such as filaments and voids appear more prominently. On the other hand, when considering the most massive haloes (HF), their assembly is distinctly different from their lower mass counterparts. In the distribution of the most massive haloes (HF) the structures are larger and voids span greater distances than in the case of the least massive haloes (HA and HB). Simply by considering the spatial distribution of haloes in the P-Millennium, we observe how haloes of different mass ranges trace different structural patterns in the cosmic web. These structural patterns are precisely what our topological analysis aims to capture by means of the persistence diagrams and Betti curves.

The dependence of the spatial clustering of haloes on their mass has been known since the seminal study by Kaiser (1987). It led to the concept of *bias* and the realization that the spatial distribution of galaxies, haloes, and clusters that emerged from the cosmic matter field may offer a distorted view of the underlying dark matter structure. It manifests itself directly in terms of second order clustering, as measured by the two-point correlation function $\xi(r)$ of galaxies and haloes (Peebles 1980). It involves the systematic trend of the correlation function amplitude with the mass of the objects. In general, biasing entails a higher clustering amplitude for more massive objects (Kaiser 1987; Mo & White 1996, 2002; Mo, Jing & White 1997; Mo, van den Bosch & White 2010; Desjacques, Jeong & Schmidt 2018). The most outstanding case is that of the cluster distribution, where various studies have demonstrated the almost linear increasing trend of clustering amplitude with cluster mass and richness (Szalay & Schramm 1985; Bahcall, Batuski & Olowin 1988; Bahcall et al. 2003; Berlind et al. 2006; Estrada, Sefusatti & Frieman 2009).

Based on this observation, we first address the two-point clustering of the different halo subsamples before investigating the higher order topological imprint.

3.3 Clustering in P-Millennium subsamples

In our P-Millennium halo sample, the mass dependence of the halo two-point correlation function is clearly borne out by its behaviour for the different halo mass bins (see left-hand panel of Fig. 3). To a first approximation, the two-point correlation function behaves as a power law,

$$\xi(r) = \left(\frac{r}{r_0}\right)^\gamma, \tag{7}$$

so that it is fully characterized in terms of the power-law slope γ and the clustering amplitude in terms of its correlation or clustering length r_0. Table 2 (third and fourth columns) lists the parameters r_0 and γ for the P-Millennium halo subsamples.[4] There is a clear systematic trend, where more massive haloes (such as HD, HE, and

[4]Note: to compute the two-point correlation function of each halo population in the P-Millennium, we use the *Landy-Szalay* estimator (Landy & Szalay 1993; Kerscher, Szapudi & Szalay 2000), which is given by

$$\xi(r) = \frac{\langle DD \rangle - \langle 2DR \rangle + \langle RR \rangle}{\langle RR \rangle}, \tag{8}$$

where $\langle DR \rangle$ is the normalized number of pair counts between the data and the Poisson point distribution while $\langle DD \rangle$ is the normalized number of pair
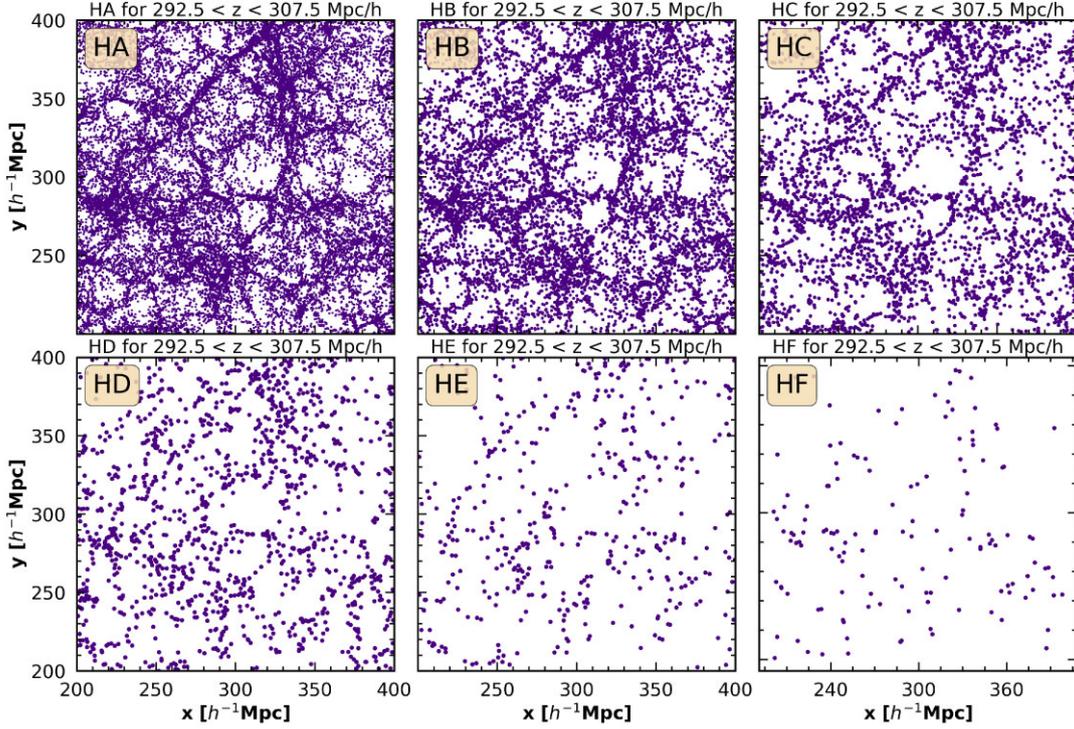
Figure 2. Spatial distribution of haloes in the P-Millennium. This plot represents the point distribution of haloes of a $200 \times 200\,h^{-1}$Mpc slice in the P-Millennium, cut through a slice of $15\,h^{-1}$Mpc thickness. From left to right in each row: HA, HB, HC, HD, HE, HF. Notice that each of the panels represents the same simulation box, but with haloes subsampled from different mass ranges (see Table 1 for the exact ranges of all halo populations). Each halo population highlights a different morphological aspect (or varying scales of connectivity) of the cosmic web by sampling a different mass range. This is also reflected in the difference in number of haloes for each halo population, with the heaviest haloes tracing the high-density nodes, and the lightest haloes weaving the whole filamentary structure.

Table 2. Clustering correlation parameters of haloes in the P-Millennium simulation. For each halo class and its logarithmic mass range, we show the respective clustering length r_0 and the exponent of the power-law fit γ. $\widetilde{r_0}$ and $\widetilde{\gamma}$ are the same parameters for the halo populations with second-order clustering normalized using the respective clustering lengths r_0.

Halo population	$\log_{10}(M)\ [h^{-1}\mathrm{M_\odot}]$	$r_0\ [h^{-1}Mpc]$	γ	$\widetilde{r_0}\ [r_0]$	$\widetilde{\gamma}$
HA	(10.5,11]	2.49 ± 0.01	-1.247 ± 0.008	1.009 ± 0.005	-1.29 ± 0.01
HB	(11,11.5]	2.71 ± 0.01	-1.298 ± 0.006	1.003 ± 0.004	-1.33 ± 0.01
HC	(11.5,12]	3.11 ± 0.01	-1.385 ± 0.004	0.995 ± 0.003	-1.41 ± 0.01
HD	(12,12.5]	3.87 ± 0.02	-1.417 ± 0.008	0.979 ± 0.003	-1.50 ± 0.01
HE	(12.5,13]	4.91 ± 0.02	-1.54 ± 0.01	0.99 ± 0.01	-1.55 ± 0.02
HF	(13,13.5]	6.88 ± 0.08	-1.51 ± 0.02	0.963 ± 0.007	-1.72 ± 0.03

HF) are more correlated and clustered than the lower mass haloes (HA, HB, and HC).

3.4 Clustering rescaling

To get a more balanced and objective impression of the higher order clustering aspects of the halo (sub)samples we compensate for two aspects of the halo distribution, the differences in clustering strength and the differences in number density of the halo subsamples.

We compensate for the stronger clustering of the more massive halo samples by rescaling the various subsamples by the corresponding clustering scales r_0. The scales, and coordinates x, y, z of the

counts between the data. Lastly, $\langle RR \rangle$ corresponds to the normalized number of pair counts between the Poisson point distribution.

haloes, are rescaled to dimensionless coordinates

$$\widetilde{x}: = \frac{x}{r_0}, \qquad \widetilde{y}: = \frac{y}{r_0}, \qquad \widetilde{z}: = \frac{z}{r_0}. \qquad (9)$$

The two-point correlation function of the rescaled halo samples is plotted in Fig. 3 (right-hand panel). As intended, the rescaled correlation function have equal clustering length $\widetilde{r}_o = 1$. They are also similar power-law functions, with almost equal slopes at small scales $\widetilde{r} < 2$. However, towards larger scales the rescaled correlation functions do clearly reveal systematic differences, with the higher mass samples displaying a stronger decline towards larger scales. Apparently, the different halo subsamples are not perfect self-similar representatives of an underlying distribution. Instead, this testifies of the significant presence of higher order clustering contributions.

Visually, the differences in clustering patterns between the halo (sub) samples reveal themselves optimally by sampling the same

Figure 3. Second-order clustering of haloes in the P-Millennium. The plots show the two-point correlation functions for each of the halo populations of the P-Millennium. The left-hand panel shows the unscaled correlations functions, each with a power-law fit (see Table 2 for the parameters). This clustering information was used to rescale the halo locations, which is shown in the right-hand panel. The two-point correlation functions are shown as continuous lines, and the power-law fits are shown as dashed lines. For each panel, we also show the (horizontal) line $\zeta(r) = 1$ (in black) and vertical lines at each halo class's clustering length. Where the clustering length r_0 differs significantly in the unscaled case, rescaling leads to values very close to $\widetilde{r_0} = 1$ (see Table 2 for the parameters). We observe a systematic trend where the massive haloes are more clustered than the lower mass haloes.

number of haloes within a given rescaled volume. This removes the influence of point density on the impression of clumping perceived by the human eye. Fig. 4 plots the rescaled spatial distribution of haloes in the six subsamples in the same rescaled volume, a slice of (rescaled) size $80 \times 80 \times 5$. Each of the halo samples contains a random selection of the same number of 200 haloes.

The overall impression is that of spatial patterns that largely resemble each other. This is a clear manifestation of the approximately self-similar character of the halo distribution as expressed by the power-law two-point correlation function. Overall, even the large-scale structure of the different halo distribution is largely similar. Nonetheless, we also discern a gradual and systematic shift in the nature of the large-scale patterns defined by the halo distributions, from a somewhat random character for the HA sample (top left-hand panel) towards a more and more structured geometric pattern in the most massive halo sample HF (bottom right-hand panel).

The subtle differences in structure at large scales, also in the rescaled halo distributions, are responsible for the differences seen in the corresponding rescaled second-order correlation functions and deviations from pure self-similarity. They testify of the presence of higher order correlations. Here, we therefore aim to identify structural patterns that go beyond the second-order probe of clustering. The challenge is to identify and quantify this in terms of measures that facilitate a direct relation with the visible web-like spatial pattern, its multiscale character and connectivity of structure in the distribution of haloes. It leads us to the investigation of the nature of these higher level structures with a topological underpinning, and a description in terms of *persistent topology*.

4 TOPOLOGICAL SCALE DEPENDENCE

In this section, we delve into the scale-dependence of the halo topology. First, we will treat the overall topology and connectivity of the cosmic web, traced through the different samplings of the six halo mass classes. This is done by looking at the total number of topological features (for each dimension) at different scales, represented by the Betti curves for varying filtration length-scales α. This allows a visualization of the global build-up of the connectivity, highlighting general characteristic length-scales. We then focus on the much more precise characterization of the length-scales (α_b, α_d) at which each of the topological features form and disappear, provided through the persistence diagrams. This will give insights into the multiscale nature of the cosmic web as traced by dark matter haloes.

Both Betti curves and persistence diagrams serve as topological tracers of the strength of the correlation between the structures haloes outline and the underlying geometric dimension. These tracers (after normalizing according to the two-point correlation function) quantify higher orders of structuring and the correlation of the halo mass with the underlying geometry. As such they pick up different clustering behaviour of light and heavy haloes, and allow statements about how it depends on the geometrically defined environment.

4.1 Alpha shapes of the halo distribution

Fig. 1 shows the two-dimensional alpha shape for the HC halo population inside a thin slice through the simulation volume. In a sequence of six panels, we see the development of alpha shapes for increasing value of its scale parameter α. The figure highlights the sensitivity of the alpha shape to the clustering of haloes and its multiscale nature, revealing the presence and scale of filamentary
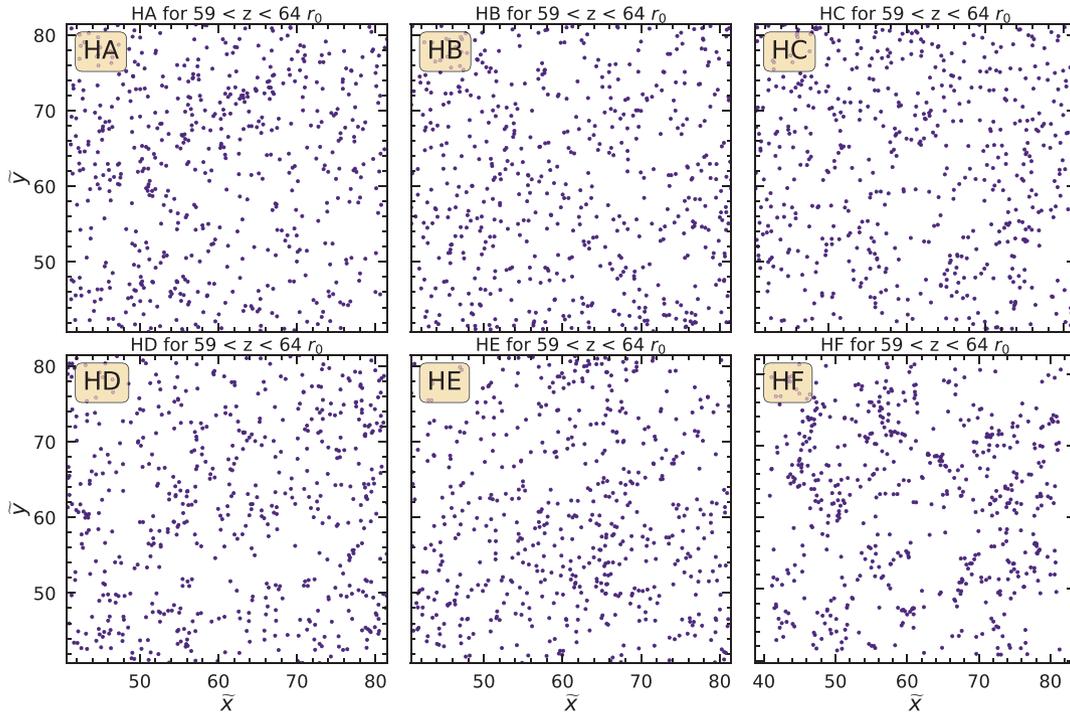
Figure 4. Rescaled halo distribution in the P-Millennium. The plot represents the point distribution of haloes of a scale-independent 80×80 (r_0) slice in the P-Millennium, cut through a scale-independent slice of $\Delta \widetilde{z} = 5$ thickness. From left to right in both rows: HA, HB, HC, HD, HE, HF. As in Fig. 2, each of the panels represents the same simulation box, but with haloes subsampled from different mass ranges (see Table 1 for the exact ranges of all halo populations) and their coordinates divided by each respective clustering length r_0 (see Table 2). Additionally, the haloes of classes HA–HE have been randomly sampled to feature the same number of haloes within the slice as appear in HF. We point out that whereas it is more difficult to distinguish between the different clustering behaviour of the mass classes, it is still possible. The heaviest haloes (HF) appear distinctly more clustered the classes of lighter haloes.

bridges, gaps or voids of haloes, and tunnels that only fill up at the highest values of *alpha*.

The vast potential and richness of the topological information on the cosmic web contained in alpha complexes becomes even more apparent when assessing the full three-dimensional situation. The two supplementary figures shown in Appendix B also testify this. They show – for three different values of α – the full three-dimensional alpha shapes for the HA halo population of the *P*-Millennium simulation, in a slice with the full simulation box width of $542.16\,h^{-1}$Mpc and $30\ h^{-1}$Mpc width, along with zoom-ins into specific regions. They reveal in meticulous detail the web-like pattern defined by the distribution of haloes, the nature of its composing structures, their mutual connections and its intricate multiscale character. Further details are outlined in the corresponding Appendix B.

4.2 Scale-independent parametrization – the sample rescaling

Following our observation in the previous section that the higher order clustering character of the halo distribution manifests itself directly in the pattern of the corresponding *rescaled* spatial point distribution, we assess its multiscale topology by analysing the *rescaled alpha shapes*. This entails the rescaling of the scale parameter α by the corresponding clustering length r_0,

$$\widetilde{\alpha} : = \frac{\alpha}{r_0}, \tag{10}$$

and with respect to the topological persistence diagrams the rescaling of the birth and death scales $\widetilde{\alpha}_b$ and $\widetilde{\alpha}_b$ of the various topological features,

$$\widetilde{\alpha}_b : = \frac{\alpha_b}{r_0}, \qquad \widetilde{\alpha}_d : = \frac{\alpha_d}{r_0} . \tag{11}$$

In the case of a purely self-similar spatial pattern, fully specified by its two-point correlation function, we would expect its topological properties to remain unaltered for the halo subpopulations. The topology would be entirely determined by the second order clustering. However, in the presence of higher order clustering the situation will be different resulting in more complex topological behaviour. The different spatial patterns seen in the scaled halo distributions in different mass ranges testify of such topological bias. To optimize the sensitivity to these topological properties, we therefore assess the Betti curves and persistence diagrams for both the regular and the scaled halo distributions.

As we will observe in the following section, also the scaled Betti curves and persistence diagrams reveal a systematic shift from low-mass to high-mass haloes. The systematic shift shows that different halo populations (of different mass ranges) trace different structural patterns at higher order clustering, and hence reflect a topological bias.

4.3 The overall topology – Betti curves

We first investigate the global topology by analysing the Betti curves for the halo distribution in the P-Millennium. We show the number of independent topological features per dimension (β_0, β_1, β_2) for both unscaled (such that α represents a physical distance) and rescaled (with respect to r_0) cases to highlight the general multiscale build-up. To ease the visualization of these curves, the
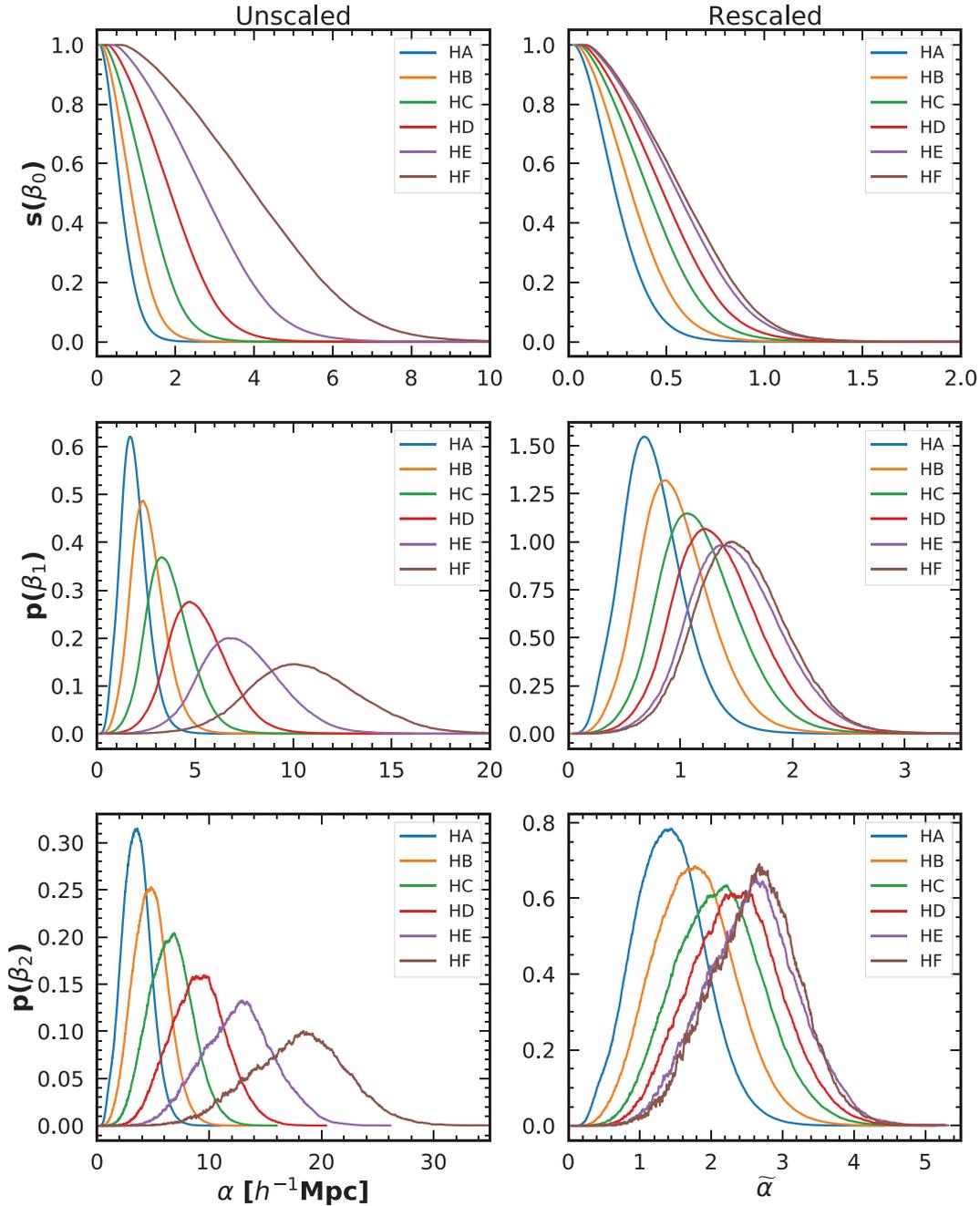
Figure 5. Betti curves for different classification of haloes in the P-Millennium. From top to bottom: standardized β_0-curve $s(\beta_0)$, corresponding to disconnected haloes; normalized β_1-curve $p(\beta_1)$, associated with filaments; and β_2-curve $p(\beta_2)$, associated with voids. The left column corresponds to the unscaled Betti curves, where α is in physical units of h^{-1}Mpc, whereas the right column corresponds to the rescaled Betti curves, where $\widetilde{\alpha} = \frac{\alpha}{r_0}$ represents a scale-independent parameter that nullifies the clustering of the two-point correlation function (see Table 2 for exact values of r_0). Note that even when we take out the two-point correlation function from the Betti curves (by considering the evolution of the alpha shape in terms of the scale-independent parameter $\widetilde{\alpha}$), the haloes still present a topology that is not self-similar with respect to each other.

β_0-curves were standardized by dividing by the number of haloes in each halo population, indicated as $s(\beta_0)$. In this case, at a value of $\alpha = 0$, the disconnected haloes correspond to $s(\beta_0) = 1$, as no connections have taken place in the alpha shape yet and the number of separate clusters of haloes is equal to the total number of haloes.

For the β_1 and β_2 curves (associated with filamentary loops and voids respectively), we carried out a normalization of the Betti curves (dividing by the area under the curve), which we denote as $p(\beta_i)$ for i

$\in \{1, 2\}$. Notice that these procedures were applied for both unscaled and rescaled Betti curves, as shown in Fig. 5.

β_0: disconnected haloes

The first row of Fig. 5 shows the β_0 curves. For the unscaled curves (top left panel), we observe a consistent trend. That is, the lower mass haloes connect up at lower values of α as the number of them (per population) is higher. More precisely, the lowest-mass haloes

(HA) have all connected at length-scale of $\alpha = 2\,h^{-1}\mathrm{Mpc}$ whereas the most massive haloes are still not fully connected at length-scales of $\alpha = 9\,h^{-1}\mathrm{Mpc}$. This indicates that it is more likely for lower mass haloes to find another such halo of the same mass range within their neighbourhood and that higher mass haloes are spread out over larger structures with connections forming for larger length-scales.

As soon as we correct the clustering length of each halo population with respect to α interesting trends start to appear. The β_0 curves start to become more similar with respect to each other in terms of their steepness and the speed with which they attain full connectivity. Since the distance at which haloes form connections is associated with their clustering length, rescaling the β_0-curves with r_0 provides a more comparable description of the topology of the different halo populations. More explicitly, we see now that the lowest-mass haloes have mostly connected at a scale-independent threshold of approximately $\widetilde{\alpha} = 0.5$, while the most massive haloes have formed most connections only at $\widetilde{\alpha} = 1.5$. This is concrete indication that the heaviest haloes trace features on larger scales than their lighter counterparts, with some relevant connections still forming even at rescaled distances exceeding the clustering length as determined from the second-order clustering. We point out this behaviour in Fig. 2 (unscaled) and Fig. 4 (rescaled).

β_1: filaments

The second row of Fig. 5 shows the β_1 curves. Again, we see a consistent trend for the unscaled β_1 curves (middle-left panel). The lower mass haloes (HA–HC) form most of their filamentary structure at values of $\alpha < 5\,h^{-1}\mathrm{Mpc}$. In the case of the most massive haloes (HE, HF), we see that most filaments appear at larger scales of $\alpha > 7.5\,h^{-1}\mathrm{Mpc}$. This indicates the birth of large-scale filaments and a highly connected (as indicated by the behaviour of the β_0-curves at the respective values of α) filamentary structure that makes up the complex network of the cosmic web. The multiscale nature of these structures is better apparent from the persistence diagrams, whereas the Betti curves serve as a straight-forward indication of overall trends and large-scale changes in the halo-topology.

Even when we correct for the clustering length of each halo population (middle-right panel), we still observe the same trend as for the unscaled case. More massive haloes (HD–HF) form filamentary loops later than the less massive ones (HA–HC). The majority of filamentary loops are formed at values of $\widetilde{\alpha} < 1$ for the three lowest mass classes, and for values between $1.5 < \widetilde{\alpha} < 2.5$ for the three classes of heavier haloes. These differences together with the broader curves for the heavy haloes (HD–HF) indicates that more massive haloes are not merely tracing a scaled-up version of the structures traced by lighter haloes, but that their connected structures form and exist over wider ranges in the probed distance space.

β_2: voids

The last row of Fig. 5 shows the β_2 curves. Similarly to the intensity persistence diagrams of dimension 2, the bottom left panel indicates that lower mass haloes (HA–HC) form most of their voids at length-scales of $\alpha < 8\,h^{-1}\mathrm{Mpc}$. More remarkable is the case of the more-massive haloes (HD–HF). For these halo populations, the β_2-curves manage to capture a rich hierarchy of void formation at length-scales of $\alpha > 10\,h^{-1}\mathrm{Mpc}$.

Although this hierarchy becomes more diffuse when we correct for the clustering length of each halo population, we still observe a difference in the formation of voids with respect to $\widetilde{\alpha}$. We observe a similar symmetry to that of filament formation around $\widetilde{\alpha} = 1$. The void structure of lower mass haloes is present at scale-independent values of $\widetilde{\alpha} \leq 2$, whereas the most significant voids only appear at scale-independent values of $\widetilde{\alpha} > 2$, corresponding to the more massive haloes.

4.4 Multiscale connectivity – persistence diagrams

Following the formalism presented in Section 2.7 and Appendix B, we computed the persistence points (α_b, α_d) and plotted α_d versus α_b. Given the large number haloes in the P-Millennium, we computed what we denote as *intensity persistence diagrams* (Pranav et al. 2017). These are equivalent to the standard persistence diagrams, except that a colour map indicates the number of counts of persistence points of a certain region on the persistence plane. This was done by constructing a 2D histogram counting, for every persistence point, the number of other pairs located in the histogram bin.

In Fig. 6, we depict the one- and two-dimensional intensity persistence diagrams of three halo populations (HA, HC, and HF). The zero-dimensional diagrams are not shown and excluded from this discussion, as all features share a birth-value α_b of zero, and they are thus less informative than in the other dimensions. We are mainly interested in studying the characterization of filaments (i.e. filamentary loops, associated with one-dimensional topological features) and voids (associated with two-dimensional topological features). The primary axes of the diagrams (bottom and left) use the rescaled values $\widetilde{\alpha}$, whereas the secondary axes (top and right) show unscaled values α. The cluster length ($\widetilde{\alpha}_{b,d} = 1$) is marked by horizontal and vertical lines. In these dimensions the persistence diagrams exhibit a characteristic, roughly triangular shape, that can be attributed to the multiscale nature of the structures under investigation.

To a large degree, their shape reflects what can be expected from a halo-based sampling of the underlying dark matter distribution. The detailed information on the process of structure formation due to gravity that can be obtained from the persistence diagram of the dark matter distribution is discussed in an earlier publication of the group (Wilding et al. 2021). Several of the relevant notions also extend to the persistence diagrams of the halo distribution, whereas there are also notable deviations.

The apex of the persistence diagram

Fig. 6 illustrates the intensity persistence diagrams. The roughly triangular shape of the diagrams is the first similar characteristic. In general, the triangular region is bounded by the diagonal and two edges converging towards a tip. For the tip of this region in particular, we coined the term *apex* of the persistence diagram (Wilding et al. 2021). At least one of the edges commonly exhibits a concave shape, in the case that both edges feature this attribute the diagram usually spouts a distinct, sharp apex.

The base of the triangular region is formed by persistence points along the diagonal. These points, with very small values of persistence, are the most short-lived topological features commonly associated with topological noise. Opposed to that, the apex consists of the points with the highest values of persistence. These points characterize the most prominent features of the halo distribution – they are visible during wide ranges of α.

The apex and its sharpness also indicate the occurrence of phase transition-like behaviours in the emerging connectivity of the cosmic web. Such a behaviour is generally found in regions where a large number of persistence points appear with similar birth- or death-values (α_b or α_d). An example of such a transition can be found
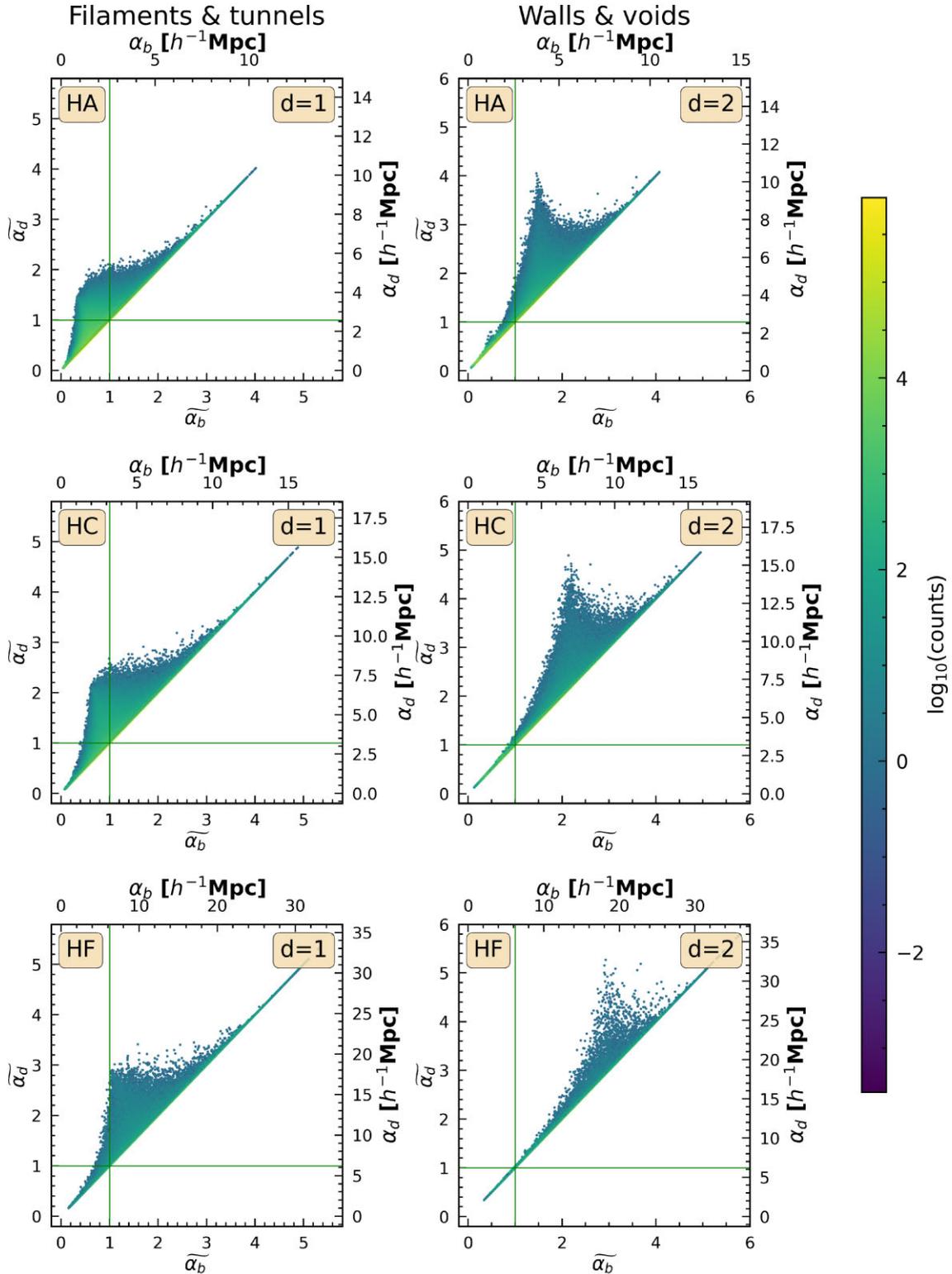
Figure 6. Persistence diagrams. We show the diagrams of dimension 1 (associated with filaments, left column) and dimension 2 (associated with voids, right column) for three different classifications of haloes in the P-Millennium. From top to bottom, we show the diagrams of haloes of the classes HA, HC, and HF (see Table 2 for the exact mass ranges). The birth and death thresholds (α_b, α_d, respectively) are given both in units of h^{-1}Mpc (secondary axes) and in the scale-independent parameters $\widetilde{\alpha}_b$ and $\widetilde{\alpha}_d$ (primary axes). The manifest differences in these birth–death diagrams between the two columns reveal the differences between haloes of a given mass in either of the two topologically defined environments. This is clear evidence for a topological bias.

in the two-dimensional persistence diagram of the lightest haloes (HA), where the apex is particularly tapered. The features in the apex indicate prominent structures that emerge within a very narrow range of birth-values α_b. The location of the apex (in terms of unscaled/rescaled birth- and death-values) will also serve as a tracer for the most prominent and distinct structures. The main variations between in the persistence diagrams shown here occur in relation to the shape and location of the apex.

Dimension 1: filaments

In the left column of Fig. 6 we show the one-dimensional intensity persistence diagrams for the halo populations HA, HC, and HF. The diagrams outline the multiscale filamentary structure traced by the dark matter haloes, separate for each population. From the diagram it is possible to analyse the emergence (birth), disappearance (death), scale, and prominence of topological features, which in this case are independent, closed filamentary loops. Immediately apparent is the flattened or rounded apex of the diagram for the lightest haloes (HA), which makes it difficult to pinpoint exact associated birth- or death-values. For the lightest haloes the most persistent features are born in a small range slightly below $\widetilde{\alpha}_b \approx 0.5$ and die at values below $\widetilde{\alpha}_d \approx 2.0$.

Moving to higher mass haloes leads to a clear sharpening of the apex, and the concentration of points with high persistence in more narrow ranges of $\widetilde{\alpha}_b$ and $\widetilde{\alpha}_d$. For the HC halo population, the slightly more narrow apex is located at birth-values in the range of $\widetilde{\alpha}_b \simeq 0.5$ to 1.0, with death-values around $\widetilde{\alpha}_b \simeq 2.3$.

For the heaviest haloes the characteristic length-scales of the prominent filamentary features shifts to even higher densities, with the apex now converging into a sharp tip. This occurs at a birth-value of $\widetilde{\alpha}_b = 1$, and at a death-value of slightly below $\widetilde{\alpha}_d = 3$.

The changes in the shape of the apex are a clear indication that when considering a higher mass class of haloes, on the one hand prominent features traced by those haloes form connections at more similar length-scales. On the other hand, the number of structures of comparable persistence (i.e. with similar distance to the diagonal) decreases, going from a larger number of features (in a flat, plateau-like apex) forming well below the respective clustering length (HA) to much fewer features forming at almost exactly the clustering length (HF).

Naturally, very massive haloes reside in massive filaments. But the clustering length, as a characteristic probe of the degree of clustering (and thus mass) of a halo population, together with its relation to the existence and location of topological features of high persistence (i.e. massive loops of filaments) allows deeper insights. In particular, the differences in the location of high-persistence features when the halo mass changes allow the identification of higher order trends between halo mass and the underlying filamentary network.

Dimension 2: voids

The right-hand column of Fig. 6 shows the two-dimensional intensity persistence diagrams for the halo populations HA, HC, and HF. These diagrams trace the multiscale nature of the void structure as woven by dark matter haloes.

Whereas the one-dimensional diagrams (tracing the filamentary components of the halo structure) shown in the left-hand column feature three very differently shaped apexes, the shapes of the two-dimensional diagrams' apexes are largely similar to each other. They differ mostly in their exact location, and in their prominence, determined by the number of points they consist of, which is mainly

due to the significantly varying numbers of haloes in the different mass classes (see Table 1).

The apexes exhibit a clear shift in location, clearer for the one-dimensional diagrams. Regarding the lowest-mass haloes (HA), we observe that the most persisting voids are formed at birth-values of $\widetilde{\alpha}_b = 1.5$ and filled at $\widetilde{\alpha}_b = 3$ to 4. Interestingly, we observe the appearance of a small peak in persistence at values of $(\widetilde{\alpha}_b, \widetilde{\alpha}_d) \simeq (0.75, 0.75)$. This likely corresponds to a small collections of low-mass haloes connecting up to form local voids at very small scales. However, these do not represent the cosmic vast voids that we are interested in quantifying in this study.

Moving to more massive haloes (HC), we see that the apex has shifted to higher birth- and death-values. Most of the void formation now occurs at a birth-value of $\widetilde{\alpha}_b \simeq 2$, and most of these voids are filled at values between $\widetilde{\alpha}_d = 4$ to 5.

For the case of the most massive haloes (HF), most of the void formation occurs at a higher birth threshold of $\widetilde{\alpha}_b \simeq 3$. Moreover, these large voids persist to death thresholds of $\widetilde{\alpha}_d \simeq 5.25$. These topological features correspond to voids of persistence values of approximately $\pi \approx 15\,h^{-1}\mathrm{Mpc}$, and indeed seem to be in accordance to the spatial distribution of haloes as shown in Fig. 2.

4.5 Distribution of persistence – lifetime curves

Finally, another different perspective for analysing persistence diagrams can be achieved with lifetime curves. These summary curves focus on a different aspect of persistence than the Betti curves, namely the persistence of a given feature. Each persistence value corresponds to a specific difference in birth- and death value, thus this style of depiction allows to analyse the multiscale nature of the halo population with a focus on the stability of features and their distribution. In Fig. 7, we show the relative prominence of features with a specific persistence by normalizing the number of features at a specific persistence. The left column uses persistence based on the distance in physical units of $h^{-1}\mathrm{Mpc}$, whereas in the right column the unscaled persistence (in units of r_0, exact value in Table 2) is used, correcting for clustering according to the two-point correlation function. The rows show the behaviour of a different dimension of features, from top to bottom these are clusters of haloes, filaments, and voids. In all panels we focus on the higher ranges of persistence (i.e. longer-lived features above $0.1\,h^{-1}\mathrm{Mpc}/0.01 \cdot r_0$).

The unscaled persistence curves of the left column immediately highlight that features with higher persistence in all dimensions are more directly associated with heavier haloes, as indicated by the sharp cut-off increasing to a higher persistence threshold with higher halo mass. This both shows the more clustered nature of heavier haloes, but also their generally lower total number. The difference in cut-off is also notably stronger for zero- and one-dimensional features, and would thus be visible in the halo clusters and the filamentary structure outlined by the different halo classes. The behaviour of the persistence curves is perceptibly (and unsurprisingly) more similar once we take clustering into account. The sharp cut-offs now occur at very similar values slightly above a persistence value of 1, and in particular the two-dimensional persistence curves start to exhibit almost self-similar behaviour when rescaled. However, some deviations between the mass classes of haloes still remain, and the general trend of the unscaled curves is unchanged, with lower mass haloes having fewer prominent high persistence features.

To conclude, methods of persistent homology tell us that high persistence features trace the most prominent components, usually spanned by the most massive haloes. Analysing the persistence of the cosmic web is possible on three layers of detail, starting from
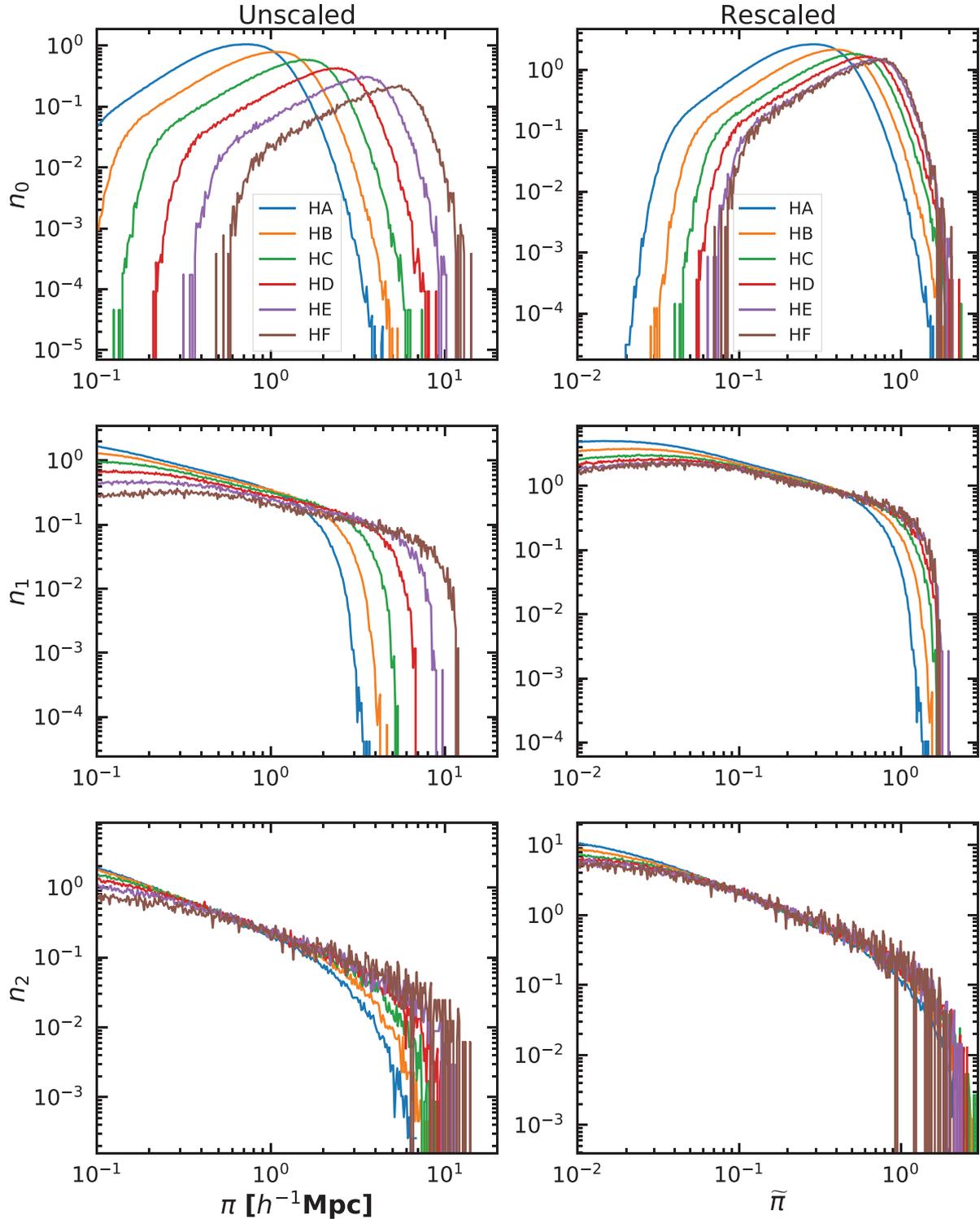
Figure 7. Persistence lifetime curves for all halo populations in the P-Millennium. The curves show the fraction of persistence points n_d at a specific persistence value (as defined in equation 6). From top to bottom, we show the curves corresponding to disconnected haloes, filaments, and voids. The left column corresponds to the unscaled persistence curves, where α is in physical units of h^{-1} Mpc, whereas the right column corresponds to the rescaled lifetime curves in units of r_0 (see Table 2 for exact values of r_0). Note that even when we take out the two-point correlation function from the persistence date (by considering the evolution of the alpha shape in terms of the scale-independent parameter $\tilde{\alpha} = \frac{\alpha}{r_0}$), the haloes have a slightly different persistence. More remarkably, we see that the filaments (middle row) have a much sharper cut-off than the voids (bottom row). This is reflected already in the intensity persistence diagrams (see Fig. 6), the voids have a more gradual decrease in high persistence features, whereas the filaments have a much more clear break in the most persistent features.
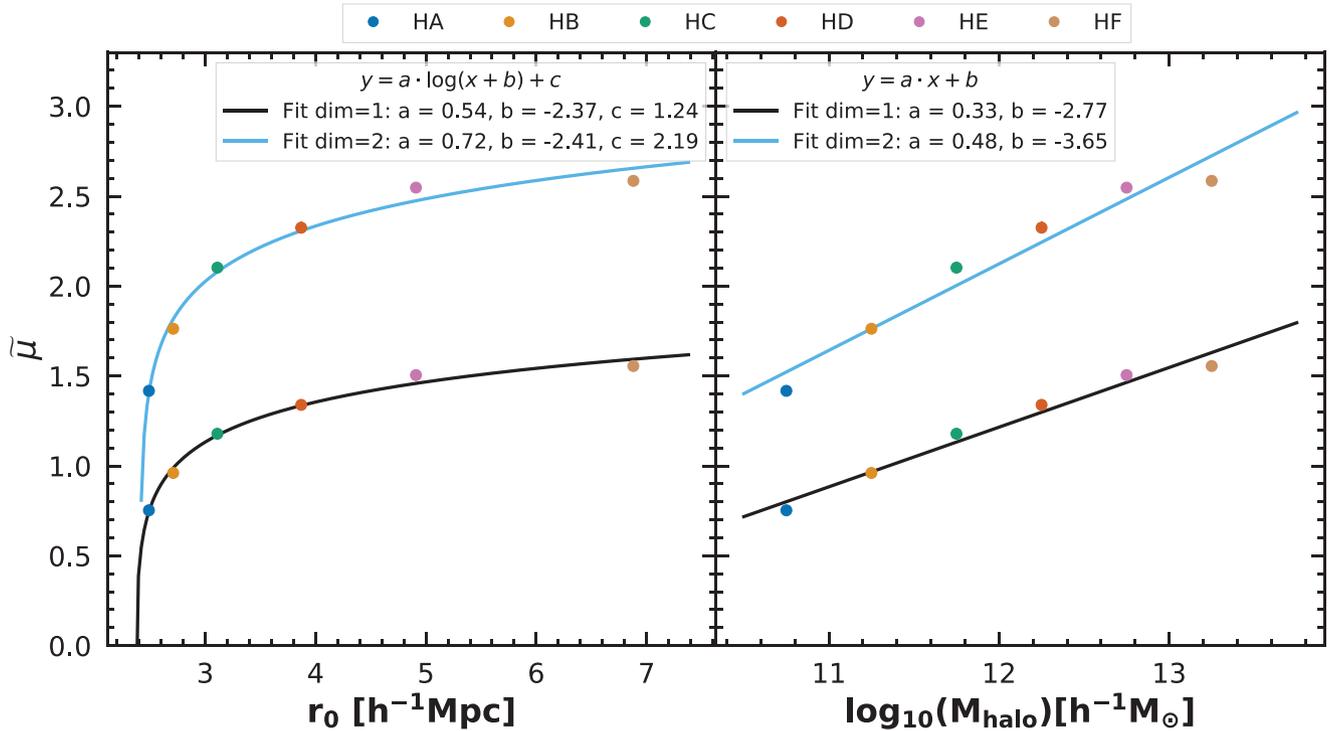
Figure 8. Betti curve scaling: dependence Betti mean on P-Millennium halo population. We show the mean μ of the Betti curves (in each respective dimension) against the clustering length r_0 (left panel) and the halo bin mass (right panel). We model the behaviour of the Betti curves' modes depending on the reference frame by a natural logarithmic or a linear model (indicated in the legend).

Betti curves, the two-dimensional intensity persistence diagrams, and ending with persistence curves. Where before we used these methods to directly investigate the structure found in different mass classes of dark matter haloes, we now focus on the systematic trends that were observable. We will quantify these trends and highlight the associated topological bias.

4.6 Global topology: Betti curves

First, we quantify the location of the maxima of the Betti curves (see Fig. 5) – and thus any displacement and corresponding change of the characteristic length-scales it encodes – through its mean, which is a good tracer of the location of the maximum of the curve (i.e. the mode). To obtain the mean we fit a skew normal distribution (O'hagan & Leonard 1976; Azzalini 1985) to the Betti curves. We use this distribution to represent Betti curves also in earlier works (see, e.g. Pranav 2015), and in particular to parametrize the evolution of the cosmic web topology in a ΛCDM universe (Wilding et al. 2021). It is defined as

$$f(x|\xi, \omega, \alpha, c) = \frac{2c}{\omega} \phi \left(\frac{x - \xi}{\omega} \right) \Phi \left(\alpha \frac{x - \xi}{\omega} \right), \qquad (12)$$

where the function $\phi(x)$ is the unit variance zero mean normal probability density distribution defined on the real line, and $\Phi(x)$ is its cumulative distribution. The variable x corresponds to the filtration value (e.g. α), ξ is a location parameter, ω a shape parameter, and c is a normalization constant. The mean μ can be obtained from the above fitting parameters through the relation

$$\mu = \xi + \omega \frac{2}{\pi} \delta, \qquad \delta = \frac{\alpha}{\sqrt{1 + \alpha^2}}. \qquad (13)$$

Table 3. Topological bias: linear relation between the logarithmic halo mass and mean of the rescaled Betti curve ($\widetilde{\mu}$, first and second rows, see Fig. 8), and the apex (rescaled birth value $\widetilde{\alpha}_b$, third and fourth rows, see Fig. 9).

		a	b
Betti curve ($\widetilde{\mu}$)	Filamentary structure	0.33 ± 0.03	-2.8 ± 0.3
	Void structure	0.48 ± 0.05	-3.6 ± 0.6
Apex ($\widetilde{\alpha}_b$)	Filamentary structure	0.15 ± 0.01	-1.0 ± 0.1
	Void structure	0.48 ± 0.06	-3.5 ± 0.7

In Fig. 8, we show the influence of the halo mass influences on the mean of the scaled Betti curve, thus corrected for clustering. To investigate the relation to the halo mass itself, we show the change of the mean both depending on the clustering length r_0 of a halo mass class, as well as the relation to the halo mass itself. Depending on this reference frame we model the behaviour using either a logarithmic or a linear function. The parameters of the fits are given in the legend of Fig. 8. The parameter μ including uncertainties is given in Table 3.

In both cases we observe clear trends between the scaled mean and the mass of a specific halo bin. Haloes in particular mass ranges form structures and features with an underlying topology that is not simply a differently scaled version of haloes with a lower or higher mass. Instead, we see a distinctly different (topological) character, that is, a difference in prominence of one- and two-dimensional topological features depending on the mass range. This effect is apparent through the one- and two-dimensional trends differing in their slope, which points to a stronger correlation of halo mass and the two-dimensional topological features. These structures – haloes in walls that surround voids – experience a stronger change when looking at different halo masses than the one-dimensional filamentary structures.

4.7 Multiscale topology: persistence diagrams

We identify a second topological tracer through the persistence diagrams by looking at the location of the high-persistence apex. This region characterizes long-lived, stable and prominent topological features, and is less prone to influence from short-lived and noisy features. The apex represents a highly relevant transition region in the topology of the halo population. For a sharp and narrow apex, e.g. visible in the two-dimensional persistence diagrams in Fig. 6, this means the persistent structures are created at similar length-scales, i.e. within a small range of values of α_b. The apex thus represents a meaningful point for the change in the topology of these haloes, as features formed at both lower and higher values of α_b have a shorter persistence. Tracing its location (across different dimensions and halo populations) therefore allows to quantify these persisting structures themselves.

For each halo population, we compute α_b by averaging over the 25 pairs with the highest persistence $\pi = \alpha_d - \alpha_b$. In Fig. 9, we show the dependence of the apexes' birth value (as a quantifier for the location) on the mass of the halo bin. The top row uses the unscaled (not corrected for clustering) location. It indicates the connection length α in (comoving Megaparsec) at which the most persistent topological features of halo-structure form. The exact behaviour differs for the dimension in question. As expected, the formation of a filamentary halo structure (one-dimensional, in blue) consistently occurs at a lower connection length than the formation of a void-like halo-structure (two-dimensional, in grey), as the density in filaments is higher than in voids, and the distances between haloes shorter. When considering the location of the unscaled apex, the trends are well represented either by a linear fit when dependent on the clustering length (Fig. 9, top left), or by an exponential fit when dependent on halo mass (top right). The observed trends are similar to the ones for the mean of the Betti curves: a higher halo mass (and thus a higher clustering length) corresponds to a higher birth scale of the apex. The effect is again stronger for haloes weaving the void-like structure.

When we normalize for clustering, the influence of the halo mass on the detected topology is slightly reduced, but clear trends are still discernible (Fig. 9, bottom row). Where we previously used a linear and an exponential model, we now use a logarithmic model for the relation to the clustering length (bottom left), and a linear model for the relation to the halo mass (bottom right). In this representation, the similarity to the mean of the Betti curves as a topological tracer is particularly clear. Once normalizing second-order clustering, we are able to identify structure beyond the two-point correlation function and quantify it using topological tracers.

Focusing on halo mass as a discriminating attribute, we see that for most halo masses, the filamentary structure is formed through connections with a length of below one clustering length (between 0.5 and 1 r_0 for HA–HE, and slightly above 1 for HF). We note that the behaviour of the filamentary structure when expressed through the location of the apex is different compared to the mean of the Betti curves, and likely more accurate due to less influence of short-lived features. The void structure forms over a wider range of connection lengths, similar to what is apparent from the Betti curves (between 1.5 and 3 r_0).

4.8 Topological bias: quantitative relationships

To summarize, both topological tracers we introduce above quantify a topological bias we observe in the correlation between the mass of dark matter haloes and their geometrical environment. This correlation can be identified through the influence that a change

in the mass of the haloes used to sample the cosmic web has on the persistent, multiscale structure. The persistent structure is represented through persistent diagrams, in which we also pick up any change to topological build-up (e.g. in the connectivity) caused by a different sampling of the cosmic web. For a structure that features only second-order clustering, i.e. that is completely characterized by its two-point correlation function, but where clustering length still changes with the mass class, rescaling with the clustering length will remove any mass dependence of clustering. In this case, we would also not expect to pick up any change of the topological characteristics, i.e. we would expect a constant relation for the all curves shown in Figs 8 and 9.

Since we do pick up clear trends, the clustering that is intrinsic to the halo distribution is different not only between different mass classes, but also exhibits distinct higher orders of clustering. The appearance of these higher orders of clustering is of course not surprising – quite the contrary. However, the use of topological methods offers a novel approach at quantifying the complex clustering within the large-scale structure, different from using, for example, the three- or four-point correlation function. Instead of adding more points which can correlate, we include the rich topological information to determine the relation between the nature of clustering and an intrinsic structural dimension of the features under investigation. This allows us to observe that the mass has not the same influence on one-dimensional topological features (filamentary loops) as on two-dimensional features (voids). Instead we observe that the change in topology is stronger (the slope in Figs 8 and 9 is higher) for the latter – the mass of a dark matter halo has a larger influence on its embedding in the void structure. Haloes of low to medium mass are more evenly distributed in the filamentary structure irrespective of the prominence (persistence) of features, whereas heavy haloes are bound stronger to prominent features in the void structure.

5 SUMMARY AND CONCLUSIONS

This study investigates the multiscale topology of the spatial web-like pattern of the cosmic web, in an ΛCDM cosmological context, as traced by the population of dark matter haloes. This involves the question in how far the filaments and voids seen in the halo and galaxy distribution is representative for those present in the underlying dark matter web-like network.

Betti numbers quantify the number and prominence of voids, filaments and tunnels, and supercluster complexes in the matter distribution, and reflect the connectivity of these topological features. The multiscale structure and connectivity of the various topological features is studied by means of persistence, in which one follows the scale at which individual topological features emerge, and the scale at which such features disappear as they merge with other features. The systematic inventory of these events, in the form of the different dimensional persistence diagrams, offers a direct view on the multiscale nature of the cosmic mass distribution, its various morphological constituents and the nature of the hierarchical build-up of structure in the Universe.

In a range of previous publications we developed the use of homology and persistence to the analysis of the primordial matter distribution and the cosmic web (Van de Weygaert et al. 2011; Pranav et al. 2017; Feldbrugge et al. 2019; Wilding et al. 2021). In Wilding et al. (2021) we followed the topological evolution of the cosmic web in ΛCDM simulations, and established the close link between topological structure and key transitions in the cosmic structure formation process. The study provides a rich representation of the intricate and complex nature of the multiscale web-like network in
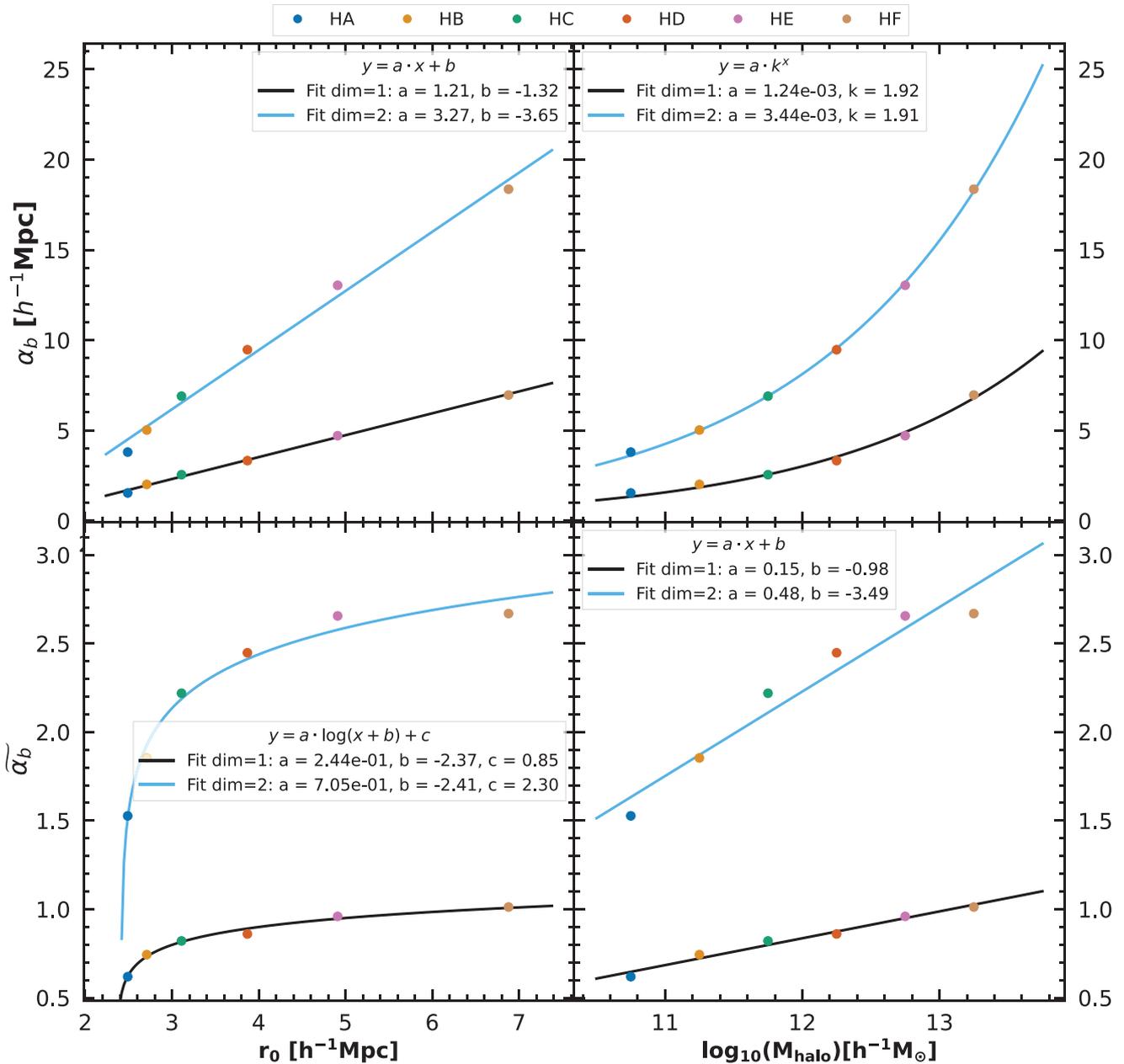
Figure 9. Apex of the intensity persistence diagrams in the P-Millennium. Here, the apex of the triangle-like shape in the intensity persistence diagrams (as shown in Fig. 5), characterized through its birth value, is plotted as a function of the clustering length of each halo populations (left column), as well as the bin mass (right column). We show the unscaled apex in the top row and the scaled apex in the bottom row. The birth value α_b of the apex represents a turning point in the topology of different halo populations, both for filaments (dimension 1) and voids (dimension 2). Depending on the reference frame, we use linear, logarithmic, or power-law models (indicated in the legend) as bases to determine fitting parameters.

which dark matter has organized itself as a result of the gravitational evolution of the primordial density and velocity field.

Here, we investigate the systematic dependence of the topology of the corresponding spatial halo distribution on the intrinsic properties of the haloes, specifically that of the mass of the haloes. The intention is to explore in how far the discrete sparse sample of haloes, and implicitly that of observed galaxies in galaxy surveys, reflects the intricate web-like structural pattern of the underlying dark matter distribution. In other words, in how far the topological structure found in the halo or galaxy population represents that of the dark matter distribution, and in how far is it significantly different. The presented topological analysis addresses the homology of the dark

halo distribution in terms of the Betti numbers as a function of spatial scale, and of the corresponding persistence characteristics.

5.1 Halo sample

The population of dark haloes in this study stems from the Planck-Millennium (Baugh et al. 2019) (also see Ganeshaiah Veena et al. (2018)), a state-of-the-art dark matter particle simulation in a periodic box of side length $542.16\,h^{-1}$Mpc. In this study, we consider haloes with masses in excess of $3.2 \times 10^{10}\,h^{-1}\mathrm{M}_\odot$, involving a sample of 1.13×10^7 haloes.

To assess the halo mass dependence of the topology and persistence of the P-Millennium halo sample, we extract six mass-defined halo subsamples. The subsample with the lowest mass haloes is that of haloes with mass between $10^{10.5} - 10^{11}\,h^{-1}\mathrm{M}_\odot$, the one with the highest mass haloes comprises haloes in the mass range $10^{13.0} - 10^{13.5}\,h^{-1}\mathrm{M}_\odot$. For each of the subsamples, we carry out a full computational Betti number and persistent homology analysis.

5.2 Alpha shapes

A key aspect of our homology analysis is that we study the discrete spatial halo distribution *directly* as a function of *spatial scale*, translating into a computational scheme based on *alpha shapes* (Edelsbrunner, Kirkpatrick & Seidel 1983; Edelsbrunner & Mücke 1994). These *simplicial complexes* – consisting of cells, faces, edges and vertices – are subsets of the Delaunay tessellation defined by the scale α.

The principal purpose of our study is to assess the structural scale dependence of (super)clusters, filaments and their associated tunnels, and voids. While most previous topology studies involved the dark matter *density field*, to avoid the intermediate step of inferring the corresponding density field we work directly via the *distance field* defined by the halo samples. For a discrete point sample it is in fact most efficient to base the computational analysis on a *simplicial complex* that is topologically equivalent to the corresponding continuous distance field (see, e.g. Edelsbrunner & Harer 2010; Boissonnat, Chazal & yvinec 2018; Carlsson & Vejdemo-Johansson 2021). To this end we invoke the Delaunay tessellation generated by the spatial halo distribution (Delaunay 1934; Van de Weygaert 1994; Okabe et al. 2000) to offer a discrete representation of the distance field.

Probing the full multiscale topology of the halo distribution involves the study of all individual changes in topology over an entire range of spatial scales. It naturally leads to the concept of *distance field filtrations*, and within its simplicial representation to that of *alpha shapes*. The filtration of a field is a subset of the field (manifold) that obeys a specific condition: usually it is the superlevel or sublevel set defined by a certain threshold. Applied to the Delaunay tessellation, the α *shape* is a simplicial filtration defined by a scale α, whose simplicial elements are the Delaunay simplices that represent connections up to the distance threshold α. Changes in the topology occur when larger Delaunay simplices get incorporated as the distance threshold increases.

5.3 Topology of the halo-sampled cosmic web

The topological analysis of the six mass-based halo samples reveals a clear mass dependence. For all three Betti number curves $\beta_0(\alpha)$, $\beta_1(\alpha)$, and $\beta_2(\alpha)$ we find a stark and systematic difference between the different halo samples. As the halo mass increases, we see the Betti curves become broader and extend out to larger scales α. We also see that the amplitude of the β_1 and β_2 curves decrease systematically as the halo mass increases.

The systematic shift in characteristic topological scale is a direct consequence of the spatial structures delineated by the haloes in the corresponding mass range. The higher mass haloes delineate voids and filaments of a larger scale than those traced by the lower mass haloes. The characteristic scale of filaments shifts from around 2 for the lowest halo mass range to around $10\,h^{-1}\mathrm{Mpc}$ for the highest mass haloes. Voids delineated by the halo distribution have a spatial scale that is roughly twice as large, ranging from 4 for the low halo mass range to $20\,h^{-1}\mathrm{Mpc}$ for the high-mass haloes. The increasing size of the filaments and voids goes along with a lowering of their total

number. In other words, high-mass haloes trace the fewer and larger specimen of these structural features.

This is, as expected, also mirrored in similar scale shifts in the persistence diagrams. Overall, the topological features traced by high-mass haloes are born and disappear at larger scales. Particularly outstanding are the persisting features in the diagrams. The apex in the β_0 diagrams, the sharp maximum that marks the scale at which the most dominant voids emerge, also sees a similar shift to larger scales as the halo mass increases. The immediate implication is that most voids in higher halo mass samples emerge at systematically larger scales.

5.4 Sampling and clustering

The fact that the higher mass haloes only trace larger filaments and voids, while missing out on the smaller and more tenuous ones, is not unexpected.

One factor is that of the lower abundance of higher mass haloes. As a result of the larger mean halo distance of the high-mass sample, they are not able to trace small filaments and voids. In other words, the high-mass halo samples offer a sparser and more diluted view of the overall structure. A physically more significant effect is that of the clustering of the various halo samples. It is known that higher mass haloes are intrinsically more strongly clustered (Kaiser 1984). The stronger clustering translates into a more clumpy distribution. By default, this will entail a pattern marked by large-scale features, such as larger underdense regions and hence larger voids.

To first approximation, the strength and scale of clustering is encapsulated in the second-order – two-point – correlation function. However, complex structural patterns such as those seen in the cosmic web and which here we seek to describe topologically entail higher order contributions. Given our interest in significant differences in the intrinsic topology of the web-like patterns traced by the different halo samples, we therefore need to compensate for the clumpiness and scale dependence implied by the two-point correlation function of the halo samples.

The two-point correlation functions of the halo samples are simple power-laws, whose sole difference is that of the specified amplitude in terms of the clustering length r_0. Hence, to remove the effect of the differences in two-point correlation functions between the different halo samples we renormalize the scales α in each halo sample by the corresponding clustering length, $\tilde{\alpha} = \alpha/r_0$.

A major asset of the rescaling is that it enables the comparison on an equal footing of the spatial patterns outlined by the halo populations. Differences in rescaled spatial structure are manifestations of the presence of higher order correlations: the two-point correlation function does not entail any phase correlations and is therefore insensitive to the presence of complex spatial patterns in the distribution of haloes (Coles & Chiang 2000; Coles 2009). It is the rich topological language of (persistent) homology that offers a visually and physically accessible characterization of these higher order influences.

5.5 Topological bias: definition

The major finding of our analysis is that stark and systematic mass dependent differences in topological structure remain between the renormalized halo samples. If anything, they are even more outstanding, following strict linear relations between topological quantities and (logarithmic) halo mass. The significant implication is that the web-like patterns and structures, such as filaments and voids, probed by haloes of different mass differ fundamentally.

For the rescaled halo samples, we observe that the Betti number curves $\beta_0(\tilde{\alpha})$, $\beta_1(\tilde{\alpha})$, and $\beta_2(\tilde{\alpha})$ (see Fig. 5) systematically shift towards higher (renormalized) scales as a function of halo mass:

(i) Light haloes, less than $10^{12} M_\odot$, arrange themselves along filamentary structure on scales *less than* the clustering length.

(ii) Higher mass haloes trace filaments on scales larger than r_0. This also holds for voids, for haloes over the entire mass range.

(iii) However, the rescaled Betti curves hardly broaden, while the amplitude of the $\beta_1(\tilde{\alpha})$ curves for filaments and tunnels only decreases slightly, and remains roughly similar for the void $\beta_2(\tilde{\alpha})$ curves.

(iv) Hence, in terms or renormalized scale: with increasing halo mass we hardly see a change in the number of voids, along with a moderate decrease in the number of filaments.

The strong nature of the systematic topological differences between filaments and voids traced by low-mass haloes and those traced by high-mass haloes is underlined by unequivocal linear quantitative trends (see Figs 8 and 9). We observe a linear relation depending on the logarithmic halo mass:

(i) The mean (renormalized) spatial scale of the Betti curves, marking the average scale of the corresponding topological features.

(ii) The (renormalized) spatial scale at the apex of the persistence diagrams for $d = 1$ (filaments) and $d = 2$ (voids). The apex of the persistence diagram marks the emergence (birth) of topological features, and encodes the hierarchical evolution that produced these.

(iii) The proportionality factor a of the linear relations for one- and two-dimensional topological features differs: the relation is steeper for the $d = 2$ situation of voids as compared to that for the $d = 1$ situation for filaments and tunnels.

These observations reveal the fundamental nature of the changing topological character of the spatial halo distributions. The immediate implication is that filaments and voids traced by haloes of different mass are fundamentally different, they differ in their topological character. Moreover, the dependence on halo mass in tracing the void population and their structure appears to be stronger than that for filaments.

The fact that the higher order structure of the cosmic web depends so strongly on the tracing halo population, and that this finds its expression in strong linear relationships of topological quantities with the (logarithmic) halo mass, may be indicated as *topological bias*. It means that haloes of different mass trace environments with different topological signature. It relates to a novel kind of halo bias, not related to second-order clustering nor directly to the underlying matter density but to the topological features in which they are located. Moreover, it affects more strongly the voids than the filaments in the cosmic web.

5.6 Topological bias: implications

Topological bias specifies that filaments and voids fundamentally differ when seen in the spatial distribution of different populations of haloes and galaxies. It is a manifestation of fundamental differences between the landscapes of filaments, tunnels, walls, and voids that constitute the structural elements of the cosmic web.

Overall, we may understand *topological bias* as the origin of the fundamental differences in structure between the filament and void population traced by high-mass versus low-mass haloes. Low-mass haloes also populate the more tenuous structures in the cosmic mass distribution. Reflecting their hierarchical build-up, the filamentary network of the cosmic web branches into smaller tendrils (see e.g.

Cautun et al. 2014). The substructure of voids – the *void hierarchy* – is the result of an even more complex hierarchical history (see e.g. Dubinski et al. 1993; Sheth & van de Weygaert 2004; Aragon-Calvo & Szalay 2013). While the low-mass haloes manage to outline the finer features of the hierarchically evolved cosmic web, high-mass haloes tend to exclusively concentrate along the largest structures in the cosmic web. A closer investigation of the origin and evolution of topological bias would shed light on the possible relation to other aspects of halo and galaxy bias, amongst which that of assembly bias (Gao, Springel & White 2005; Wechsler et al. 2006; Dalal et al. 2008; Mao, Zentner & Wechsler 2018).

As a result, low-mass haloes succeed in delineating the topological fine structure of the cosmic web. Evidently, the observational complication of finding low-mass haloes and low-luminosity galaxies will remain a major challenge in the ability to study in the observational reality the full richness of the topological structure of the cosmic web and topological bias.

Topological bias will impact on a wide range of cosmologically interesting aspects of the cosmic web. It will influence the outcome of studies of the filament population (Jones, van de Weygaert & Aragón-Calvo 2010; Tempel et al. 2014; Libeskind et al. 2018), and their use to infer intrinsic alignments between the spin of galaxies – and other galaxy properties – and the harbouring filaments (see e.g. Porciani, Dekel & Hoffman 2002a, b; Aragón-Calvo et al. 2007b; Schäfer 2009; Jones, van de Weygaert & Aragón-Calvo 2010; Codis et al. 2012; Tempel & Libeskind 2013; Alpaslan et al. 2014a; Jarrett et al. 2017; Ganeshaiah Veena et al. 2018, 2021; Kraljic et al. 2018; Welker et al. 2020). It affects even more strongly the voids outlined by haloes, and should be taken into account when seeking to derive the size distribution of cosmic voids, and use these as a measure of the underlying cosmology (see e.g. Sheth & van de Weygaert 2004; Platen, van de Weygaert & Jones 2008; Van de Weygaert & Bond 2008; Alpaslan et al. 2014b; Pisani et al. 2015, 2019; Van de Weygaert 2016; Verza et al. 2019).

In conclusion, the implications of the existence of topological bias will be particularly important for the analysis and interpretation of the spatial galaxy distribution in galaxy redshift surveys. This is true for existing surveys, such as 2dFGRS, SDSS, GAMA, 2MASS, and VIPERS (Colless et al. 2003; Tegmark et al. 2004; Driver et al. 2009; Huchra et al. 2012; Guzzo & VIPERS Team 2013) as well as for the upcoming major surveys that will map the cosmic web in unprecedented detail, such as DESI, Euclid, the Vera Rubin telescope, and SKA related surveys.

DATA AVAILABILITY

The data underlying this paper were provided by Carlos Frenk and will be shared on request to the corresponding author with his permission.

REFERENCES

Adler R., 1981, Classics in Applied Mathematics. Vol. 62: The Geometry of Random Fields. SIAM, Chichester, New York

Adler R., Taylor J., 2010, Random Fields and Geometry, Springer Monographs in Mathematics. Springer, available at: http://books.google.nl/books?id=38MDkgAACAAJ

Alpaslan M. et al., 2014a, MNRAS, 438, 177

Alpaslan M. et al., 2014b, MNRAS, 440, L106

Aragón Calvo M., 2007, PhD thesis, Univ. Groningen

Aragon-Calvo M. A., Szalay A. S., 2013, MNRAS, 428, 3409

Aragón-Calvo M. A., Jones B. J. T., van de Weygaert R., van der Hulst J. M., 2007a, A&A, 474, 315

Aragón-Calvo M. A., van de Weygaert R., Jones B. J. T., van der Hulst J. M., 2007b, ApJ, 655, L5

Aragón-Calvo M. A., van de Weygaert R., Jones B. J. T., 2010a, MNRAS, 408, 2163

Aragón-Calvo M. A., Platen E., van de Weygaert R., Szalay A. S., 2010b, ApJ, 723, 364

Azzalini A., 1985, Scand. J. Stat., 12, 171

Bahcall N. A., Batuski D. J., Olowin R. P., 1988, ApJ, 333, L13

Bahcall N. A., Dong F., Hao L., Bode P., Annis J., Gunn J. E., Schneider D. P., 2003, ApJ, 599, 814

Barrow J. D., Bhavsar S. P., Sonoda D. H., 1985, MNRAS, 216, 17

Baugh C. M. et al., 2019, MNRAS, 483, 4922

Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013, ApJ, 762, 109

Berlind A. A., Kazin E., Blanton M. R., Pueblas S., Scoccimarro R., Hogg D. W., 2006, preprint (arXiv:astro-ph/0610524)

Bernardeau F., van de Weygaert R., 1996, MNRAS, 279, 693

Bertschinger E., Jain B., 1994, ApJ, 431, 486

Bett P., Eke V., Frenk C. S., Jenkins A., Helly J., Navarro J., 2007, MNRAS, 376, 215

Biagetti M., Cole A., Shiu G., 2021, J. Cosmol. Astropart. Phys., 2021, 061

Boissonnat J.-D., Maria C., 2012, in Leah Epstein P. F., ed., Proc. 20th Annual European Symposium, Algorithms – ESA 2012. Springer, Berlin, Heidelberg, p. 731, available at:https://hal.inria.fr/hal-01108441

Boissonnat J.-D., Chazal F., yvinec M., 2018, Geometric and Topological Inference. Cambridge Univ. Press, Cambridge, available at: https://hal.inria.fr/hal-01615863

Bond J. R., Myers S. T., 1996, ApJS, 103, 1

Bond J., Kofman L., Pogosyan D., 1996, Nature, 380, 603

Bonnaire T., Aghanim N., Kuruvilla J., Decelle A., 2022, A&A, 661, A146

Borzyszkowski M., Porciani C., Romano-Díaz E., Garaldi E., 2017, MNRAS, 469, 594

Cadiou C., Pichon C., Codis S., Musso M., Pogosyan D., Dubois Y., Cardoso J. F., Prunet S., 2020, MNRAS, 496, 4787

Carlsson G., Vejdemo-Johansson M., 2021, Topological Data Analysis with Applications. Cambridge Univ. Press, Cambridge, available at: https://books.google.nl/books?id=NxFvzgEACAAJ

Cautun M. C., van de Weygaert R., 2011, Astrophysics Source Code Library, record ascl:1105.003

Cautun M., van de Weygaert R., Jones B. J. T., 2013, MNRAS, 429, 1286

Cautun M., van de Weygaert R., Jones B. J. T., Frenk C. S., 2014, MNRAS, 441, 2923

Codis S., Pichon C., Devriendt J., Slyz A., Pogosyan D., Dubois Y., Sousbie T., 2012, MNRAS, 427, 3320

Codis S., Pogosyan D., Pichon C., 2018a, MNRAS, 479, 973

Codis S., Jindal A., Chisari N. E., Vibert D., Dubois Y., Pichon C., Devriendt J., 2018b, MNRAS, 481, 4753

Colberg J. M., 2007, MNRAS, 375, 337

Coles P., 2009, in Martínez V. J., Saar E., Martínez-González E., Pons-Bordería M. J., eds, Lecture Notes in Physics, Vol. 665, Data Analysis in Cosmology. Springer, Berlin, Heidelberg, p. 493

Coles P., Chiang L.-Y., 2000, Nature, 406, 376

Colless M. et al., 2003, preprint (arXiv:astro-ph/0306581)

Dalal N., White M., Bond J. R., Shirokov A., 2008, ApJ, 687, 12

Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371

De Regt R., Apunevych S., von Ferber C., Holovatch Y., Novosyadlyj B., 2018, MNRAS, 477, 4738

Delaunay B. N., 1934, Bull. Acad. Sci. URSS, 1934, 793

Desjacques V., Jeong D., Schmidt F., 2018, Phys. Rep., 733, 1

DESI Collaboration, 2016, preprint (arXiv:1611.00036)

Dey T., Edelsbrunner H., Guha S., 1999, in Chazelle B., Goodman J., Pollack R., eds, Advances in Discrete and Computational Geometry. American Mathematical Society, Providence, Rhode Island, p. 109

Dey A. et al., 2019, AJ, 157, 168

Dressler A., 1980, ApJ, 236, 351

Driver S. P. et al., 2009, Astron. Geophys., 50, 5

Dubinski J., da Costa L. N., Goldwirth D. S., Lecar M., Piran T., 1993, ApJ, 410, 458

Edelsbrunner H., Harer J., 2010, Computational Topology – An Introduction. American Mathematical Society, Providence, Rhode Island

Edelsbrunner H., Mücke E. P., 1994, ACM Trans. Graph., 13, 43

Edelsbrunner H., Kirkpatrick D., Seidel R., 1983, IEEE Trans. Inf. Theory, 29, 551

Elbers W., van de Weygaert R., 2019, MNRAS, 486, 1523

Eldering B., 2005, Master's thesis, Topology of Galaxy Models, Univ. Groningen

Estrada J., Sefusatti E., Frieman J. A., 2009, ApJ, 692, 265

Euler L., 1758, Novi commentarii academiae scientiarum Petropolitanæ, 4, 109

Feldbrugge J., van Engelen M., van de Weygaert R., Pranav P., Vegter G., 2019, J. Cosmol. Astropart. Phys., 2019, 052

Ganeshaiah Veena P., Cautun M., van de Weygaert R., Tempel E., Jones B. J. T., Rieder S., Frenk C. S., 2018, MNRAS, 481, 414

Ganeshaiah Veena P., Cautun M., Tempel E., van de Weygaert R., Frenk C. S., 2019, MNRAS, 487, 1607

Ganeshaiah Veena P., Cautun M., van de Weygaert R., Tempel E., Frenk C. S., 2021, MNRAS, 503, 2280

Gao L., Springel V., White S. D. M., 2005, MNRAS, 363, L66

Giovanelli R., Haynes M. P., Chincarini G. L., 1986, ApJ, 300, 77

Giri S. K., Mellema G., 2021, MNRAS, 505, 1863

Goh T. et al., 2019, MNRAS, 483, 2101

Gott J. R. I., Melott A. L., Dickinson M., 1986, ApJ, 306, 341

Graham M. J., Clowes R. G., Campusano L. E., 1995, MNRAS, 275, 790

Guzzo L., VIPERS Team, 2013, The Messenger, 151, 41

Hahn O., Carollo C. M., Porciani C., Dekel A., 2007, MNRAS, 381, 41

Hahn O., Porciani C., Dekel A., Carollo C. M., 2009, MNRAS, 398, 1742

Hamilton A. J. S., Gott J. R. I., Weinberg D., 1986, ApJ, 309, 1

Hanami H., 2001, MNRAS, 327, 721

Hellwing W. A., Cautun M., van de Weygaert R., Jones B. T., 2021, Phys. Rev. D, 103, 063517

Hiraoka Y., Nakamura T., Hirata A., Escolar E. G., Matsue K., Nishiura Y., 2016, Proc. Natl. Acad. Sci., 113, 7035

Hong S., Dey A., 2015, MNRAS, 450, 1999

Hong S., Coutinho B. C., Dey A., Barabási A.-L., Vogelsberger M., Hernquist L., Gebhardt K., 2016, MNRAS, 459, 2690

Hoyle F., 1949, in Burgers J. M. , van de Hulst H. C., eds, Central Air Documents Office, Dayton, p. 195

Huchra J. P. et al., 2012, ApJS, 199, 26

Icke V., 1973, A&A, 27, 1

Icke V., van de Weygaert R., 1987, A&A, 184, 16

Jarrett T. H. et al., 2017, ApJ, 836, 182

Jiang L., Helly J. C., Cole S., Frenk C. S., 2014, MNRAS, 440, 2115

Jones B. J. T., van de Weygaert R., Aragón-Calvo M. A., 2010, MNRAS, 408, 897

Kaiser N., 1984, ApJ, 284, L9

Kaiser N., 1987, MNRAS, 227, 1

Kerscher M., Szapudi I., Szalay A. S., 2000, ApJ, 535, L13

Kimura Y., Imai K., 2017, Adv. Space Res., 60, 722

Kono K. T., Takeuchi T. T., Cooray S., Nishizawa A. J., Murakami K., 2020, preprint (arXiv:2006.02905)

Kraljic K. et al., 2018, MNRAS, 474, 547

Landy S. D., Szalay A. S., 1993, ApJ, 412, 64

Laureijs R. et al., 2011, preprint (arXiv:1110.3193)

Lee J., Pen U.-L., 2001, ApJ, 555, 106

Liang J., Edelsbrunner H., Fu P., Sudhakar P., Subramaniam S., 1998a, Proteins: Structure, Function, and Genetics, 33, 1

Liang J., Edelsbrunner H., Fu P., Sudhakar P., Subramaniam S., 1998b, Proteins: Structure, Function, and Genetics, 33, 18

Libeskind N. I. et al., 2018, MNRAS, 473, 1195

López P., Cautun M., Paz D., Merchán M., van de Weygaert R., 2021, MNRAS, 502, 5528

Makarenko I., Shukurov A., Henderson R., Rodrigues L. F. S., Bushby P., Fletcher A., 2018, MNRAS, 475, 1843

Mao Y.-Y., Zentner A. R., Wechsler R. H., 2018, MNRAS, 474, 5143

Maria C., 2015a, GUDHI User and Reference Manual. GUDHI Editorial Board, available at: http://gudhi.gforge.inria.fr/doc/latest/group_simplex_tree.html

Maria C., 2015b, GUDHI User and Reference Manual. GUDHI Editorial Board, available at: http://gudhi.gforge.inria.fr/doc/latest/group_persistent_cohomology.html

Mecke K. R., Buchert T., Wagner H., 1994, A&A, 288, 697

Mo H. J., White S. D. M., 1996, MNRAS, 282, 347

Mo H. J., White S. D. M., 2002, MNRAS, 336, 112

Mo H. J., Jing Y. P., White S. D. M., 1997, MNRAS, 284, 189

Mo H., van den Bosch F. C., White S., 2010, Galaxy Formation and Evolution. Cambridge Univ. Press, Cambridge

Munkres J., 1984, Elements of Algebraic Topology. Advanced Book Program, Addison-Wesley, available at: https://books.google.nl/books?id=tsIPAQAAMAAJ

Musso M., Cadiou C., Pichon C., Codis S., Kraljic K., Dubois Y., 2018, MNRAS, 476, 4877

Naidoo K., Whiteway L., Massara E., Gualdi D., Lahav O., Viel M., Gil-Marín H., Font-Ribera A., 2020, MNRAS, 491, 1709

Nevenzeel K., 2013, Master's thesis, Triangularing the Darkness, Univ. Groningen

Neyrinck M. C., Gnedin N. Y., Hamilton A. J. S., 2005, MNRAS, 356, 1222

O'hagan A., Leonard T., 1976, Biometrika, 63, 201

Okabe A., Boots B., Sugihara K., Chiu S. N., 2000, Spatial Tessellations: Concepts and Applications of Voronoi Diagrams, 2nd ed. Series in Probability and Statistics. John Wiley and Sons, Inc., Chichester, West Sussex

Paranjape A., Hahn O., Sheth R. K., 2018, MNRAS, 476, 5442

Park C. et al., 2013, J. Korean Astron. Soc., 46, 125

Peebles P. J. E., 1969, ApJ, 155, 393

Peebles P. J. E., 1980, The Large-Scale Structure of the Universe. Princeton Univ. Press, Princeton, New Jersey

Petri G., Expert P., Turkheimer F., Carhart-Harris R., Nutt D., Hellyer P. J., Vaccarino F., 2014, J. R. Soc. Interface, 11, 20140873

Pisani A., Sutter P. M., Hamaus N., Alizadeh E., Biswas R., Wandelt B. D., Hirata C. M., 2015, Phys. Rev. D, 92, 083531

Pisani A. et al., 2019, BAAS, 51, 40

Planck Collaboration, 2018, A&A, 641, A6

Platen E., van de Weygaert R., Jones B. J. T., 2008, MNRAS, 387, 128

Poincaré J. H., 1892, C. R. Acad. Sci., 115, 633

Porciani C., Dekel A., Hoffman Y., 2002a, MNRAS, 332, 325

Porciani C., Dekel A., Hoffman Y., 2002b, MNRAS, 332, 339

Pranav P., 2015, PhD thesis, Persistent holes in the Universe: A hierarchical topology of the cosmic mass distribution, Univ. Groningen

Pranav P., Edelsbrunner H., van de Weygaert R., Vegter G., Kerber M., Jones B. J. T., Wintraecken M., 2017, MNRAS, 465, 4281

Pranav P. et al., 2019a, MNRAS, 485, 4167

Pranav P., Adler R. J., Buchert T., Edelsbrunner H., Jones B. J. T., Schwartzman A., Wagner H., van de Weygaert R., 2019b, A&A, 627, A163

Reimann M. W. et al., 2017, Front. Comput. Neurosci., 11, 48

Robins V., 2006, Phys. Rev. E, 74, 061107

Robins V., 2015, in Grinfeld M., ed., Mathematical Tools for Physicists, Encyclopedia of Applied Physics. Wiley-VCH Verlag GmbH, UK, p. 211

Romano-Díaz E., van de Weygaert R., 2007, MNRAS, 382, 2

Rote G., Vegter G., 2006, in Boissonnat J.-D., Teillaud M., eds, Effective Computational Geometry for Curves and Surfaces. Springer, Berlin, Heidelberg, p. 277

Rouvreau V., 2015, GUDHI User and Reference Manual. GUDHI Editorial Board, available at: http://gudhi.gforge.inria.fr/doc/latest/group_alpha_complex.html

Sahni V., Sathyaprakash B. S., Shandarin S. F., 1998, ApJ, 495, L5

Salcedo A. N. et al., 2020, MNRAS, 491, 3061

Van de Sande J. et al., 2021, MNRAS, 508, 2307

Schaap W., 2007, PhD thesis, Univ. Groningen

Schaap W. E., van de Weygaert R., 2000, A&A, 363, L29

Schäfer B. M., 2009, Int. J. Mod. Phys. D, 18, 173

Schmalzing J., Buchert T., 1997, ApJ, 482, L1+

Schmalzing J., Gorski K. M., 1998, MNRAS, 297, 355

Schmalzing J., Buchert T., Melott A. L., Sahni V., Sathyaprakash B. S., Shandarin S. F., 1999, ApJ, 526, 568

Sevilla-Noarbe I. et al., 2021, ApJS, 254, 24

Shandarin S. F., Zeldovich Y. B., 1989, Rev. Mod. Phys., 61, 185

Shen J., Abel T., Mo H. J., Sheth R. K., 2006, ApJ, 645, 783

Sheth R. K., Tormen G., 2002, MNRAS, 329, 61

Sheth R. K., van de Weygaert R., 2004, MNRAS, 350, 517

Shivashankar N., Pranav P., Natarajan V., van de Weygaert R., Bos E. G. P., Rieder S., 2016, IEEE Trans. Vis. Comput. Graphics, 22, 1745

Sousbie T., 2011, MNRAS, 414, 350

Sousbie T., Pichon C., Kawahara H., 2011, MNRAS, 414, 384

Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726

Sutherland W. A., 2009, Introduction to Metric and Topological Spaces. Oxford Univ. Press, Oxford

Szalay A. S., Schramm D. N., 1985, Nature, 314, 718

Tegmark M. et al., 2004, ApJ, 606, 702

Tempel E., Libeskind N. I., 2013, ApJ, 775, L42

Tempel E., Stoica R. S., Martínez V. J., Liivamägi L. J., Castellan G., Saar E., 2014, MNRAS, 438, 3465

The CGAL Project, 2019, CGAL User and Reference Manual, 4.14 edn. CGAL Editorial Board, available at: https://doc.cgal.org/4.14/Manual/packages.html

The GUDHI Project, 2015, GUDHI User and Reference Manual. GUDHI Editorial Board, available at: http://gudhi.gforge.inria.fr/doc/latest/

Thélie E., Aubert D., Gillet N., Ocvirk P., 2022, A&A, 658, A139

Trowland H. E., Lewis G. F., Bland-Hawthorn J., 2013, ApJ, 762, 72

Tsizh M., Novosyadlyj B., Holovatch Y., Libeskind N. I., 2020, MNRAS, 495, 1311

Van de Weygaert R., 1991, Voids and the Geometry of Large Scale Structure. PhD thesis, Univ. Leiden, Leiden

Van de Weygaert R., 1994, A&A, 283, 361

Van de Weygaert R., 2016, in van de Weygaert R., Shandarin S., Saar E., Einasto J., eds, Proc. IAU Symp. 308, The Zeldovich Universe: Genesis and Growth of the Cosmic Web. Cambridge Univ. Press, Cambridge, p. 493

Van de Weygaert R., Bond J., 2008, in Plionis M., López-Cruz O., Hughes D., eds, Lecture Notes in Physics, Vol. 740, A Pan-Chromatic View of Clusters of Galaxies and the Large-Scale Structure. Springer Verlag, Berlin, p. 335

Van de Weygaert R., Schaap W., 2009, in Martínez V. J., Saar E., Martínez-González E., Pons-Bordería M.-J., eds, Lecture Notes in Physics, Vol. 665, Data Analysis in Cosmology, Springer Verlag, Berlin, p. 291

Van de Weygaert R. et al., 2011, Alpha, Betti and the Megaparsec Universe: On the Topology of the Cosmic Web. Springer, Berlin, Heidelberg, p. 60

Van de Weygaert R., Platen E., Vegter G., Eldering B., Kruithof N., 2010, in Mostafavi M. A., ed., Proc. International Symposium on Voronoi Diagrams in Science and Engineering. IEEE Computer Society, p. 224

Vegter G., 2004, in Goodman J. E., O'Rourke J., eds, Handbook of Discrete and Computational Geometry, 2nd edn. CRC Press LLC, Boca Raton, FL, p. 719

Verza G., Pisani A., Carbone C., Hamaus N., Guzzo L., 2019, J. Cosmol. Astropart. Phys., 2019, 040

Voronoi G., 1908, J. Reine Angew., 134, 198

Wang P., Kang X., 2017, MNRAS, 468, L123

Wasserman L., 2018, Annu. Rev. Stat. Appl., 5, 501

Wechsler R. H., Zentner A. R., Bullock J. S., Kravtsov A. V., Allgood B., 2006, ApJ, 652, 71

Welker C. et al., 2020, MNRAS, 491, 2864

White S. D. M., 1984, ApJ, 286, 38

White M., Cohn J. D., Smit R., 2010, MNRAS, 408, 1818

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. B., Nagai D., 2019, Astron. Comput., 27, 34

Yang X. et al., 2017, ApJ, 848, 60

Zeldovich I. B., Einasto J., Shandarin S. F., 1982, Nature, 300, 407

Zhang Y., Yang X., Faltenbacher A., Springel V., Lin W., Wang H., 2009, ApJ, 706, 747

Zhang Y., Yang X., Guo H., 2021, MNRAS, 507, 5320

Zomorodian A., Carlsson G., 2005, Discrete Comput. Geom., 33, 249

APPENDIX A: SIMPLICIAL COMPLEXES: VORONOI AND DELAUNAY TESSELLATIONS

This study focuses on the multiscale topology – persistent homology – of the halo distribution in the P-Millennium simulation. Given the discrete nature of the spatial halo distribution, we opt for a direct computation of the multiscale topology of the corresponding point distribution. This is accomplished on the basis of the *Delaunay tessellation* (Delaunay 1934; Okabe et al. 2000; Edelsbrunner & Harer 2010) generated by the point distribution. The Delaunay tessellation is a simplicial complex that is optimally sensitive to two key aspects of the cosmic web: the multiscale nature of the structure probed by the point distribution and its anisotropic morphology/geometry, in particular that of the prominent filamentary network and the connected flattened walls (see Van de Weygaert & Schaap 2009).

For the assessment of the multiscale topological structure of the cosmic web we need to invoke a well-defined scale-based filtration of the simplicial complex that represent the spatial structure of the point distribution. This leads directly to a key concept in computational topology and geometry, known as *alpha shapes* (Edelsbrunner, Kirkpatrick & Seidel 1983; Edelsbrunner & Mücke 1994; Edelsbrunner & Harer 2010). Uniquely defined for a particular point set P by the scale parameter α, alpha shapes correspond to a unique assembly of simplicial (geometric) structures that capture the shape and morphology of the point distribution.

This appendix provides definitions, details and illustrations of the related concepts from computational geometry and topology. Via illustrations of the three-dimensional alpha shapes of the web-like distribution of haloes in cosmological simulation, we demonstrate the meticulous means by which alpha shapes are probing the topological nature of complex spatial patterns.

A1 Simplicial complex

A simplicial complex is an ordered geometric assembly of faces, edges, nodes, and cells that mark a discrete spatial map of the volume containing a point set (Edelsbrunner & Harer 2010; Pranav et al. 2017). The geometric components of the complex are simplices of different dimensions: a cell is a three-dimensional simplex, a face or wall a two-dimensional simplex, an edge a one-dimensional simplex and a node a zero-dimensional simplex. A good impression of a two-dimensional simplicial complex can be obtained from Fig. 1, showing a slice through the Delaunay tessellation defined by the spatial halo distribution in the P-Millennium simulation, and a range of its alpha shape filtrations (see Appendix B). Zero-, one- and two-dimensional simplices are visible: they correspond to the vertices and edges of the particle distribution's tessellation.

Of key importance for our study is that when the simplicial complex is defined on the basis of a generating point set P, when the complex is volume-covering and reflecting the number density and geometry of P, the lengths of the edges in the complex represent a sampling of the corresponding distance field. We will exploit this with respect to the Voronoi and Delaunay tessellation of the point samples described in this study (Voronoi 1908; Delaunay 1934; Okabe et al. 2000).

A2 Delaunay and Voronoi tessellations

The first step of the computational procedure to analyse the halo topology is the simplicial representation of the halo distribution samples. The simplicial complex should represent the spatial characteristics of the point distribution. To optimally represent the multiscale nature of the spatial point distributions, as well as its organization into a network of anisotropic filamentary and flattened features, we invoke their Voronoi and Delaunay tessellations (Voronoi 1908; Delaunay 1934; Icke & van de Weygaert 1987; Van de Weygaert 1994). Of crucial importance is also that the Delaunay tessellation facilitates the definition of a transparent and uniquely defined simplicial *filtration*, that of the *alpha shapes* (Edelsbrunner, Kirkpatrick & Seidel 1983; Edelsbrunner & Mücke 1994; Edelsbrunner & Harer 2010) (see Appendix B).

The Delaunay and Voronoi tessellation are amongst the most well-known concepts in computational and stochastic geometry, uniquely defined for a spatial point distribution P (see Appendix A2 for the detailed definition). Delaunay and Voronoi tessellations have a range of properties that make them ideally suited for studies of the cosmic web in galaxy surveys and computer simulations (see eg. Icke & van de Weygaert 1987; Van de Weygaert 1991, 1994; Okabe et al. 2000; Van de Weygaert & Schaap 2009; Edelsbrunner & Harer 2010).

A2.1 Geometric definition

The Voronoi tessellation (Voronoi 1908) and Delaunay tessellation (Delaunay 1934) are amongst the most well-known simplicial complexes (see Okabe et al. 2000, for an extensive overview). They are uniquely and fully defined by a given spatial point distribution (i.e. for a generic non-degenerate point set). The Voronoi tessellation $\mathcal{V}(P)$ of a point set P divides up space into polyhedral cells (3D) or polygonal cells (2D). Each cell $\mathcal{V}_p$ allocates the part of space that is closer to a given point p than to any of the other points in the (finite) point sample $P \subseteq \mathbb{R}^3$. Formally, the Voronoi cell $\mathcal{V}_p$ of point p is defined as:

$$\mathcal{V}_p = \left\{ x \in \mathbb{R}^3 \mid ||x - p|| \leq ||x - q||, \forall\, q \in P \right\}. \tag{A1}$$

The *Delaunay tessellation* is the dual of the *Voronoi tessellation*, and the calculation of one of these tessellations immediately yields its dual. The Delaunay tessellation is the volume-filling triangulation consisting of (three-dimensional) tetrahedra whose four vertices are points of the sample P whose circumscribing sphere does not contain any of the other points. In the two-dimensional situation this concerns triangles, defined by three points whose circumscribing circle does not contain any other points of the sample P. Each vertex in the Delaunay tessellation is a nucleus of a Voronoi cell. The *edges* connect two sample points that share a common face in the Voronoi tessellation, constituting *natural neighbours*. Also, the *edges* of a Voronoi polyhedron pass through centroids of the triangular faces of a Delaunay tetrahedron, while the centroid of the Delaunay tetrahedron is one of the vertices of the polyhedral Voronoi cell.

There is a vast mathematical and computational literature on these volume-covering tessellations, and they know a wide range of applications in science and engineering. Many of the relevant aspects are treated in the volume by Okabe et al. (2000).

A2.2 Tessellations and the cosmic web: Sensitivity to multiscale and anisotropic morphology

A detailed assessment of the sensitivity of Voronoi and Delaunay tessellations to the nature of the generating point distribution (Schaap 2007; Van de Weygaert & Schaap 2009) revealed that they impressively follow the fundamental structural properties of the cosmic web. Voronoi cells are highly sensitive to the local number density of the point distribution (see Van de Weygaert 1991, 1994; Schaap & van de Weygaert 2000; Neyrinck, Gnedin & Hamilton 2005; Van de Weygaert & Schaap 2009). The dual of the Voronoi tessellation, the Delaunay tessellation, shares these properties to an even stronger extent.

The fully self-adaptive DTFE procedure exploits these properties (Schaap & van de Weygaert 2000; Schaap 2007; Van de Weygaert & Schaap 2009; Cautun & van de Weygaert 2011). It delivers an impressively accurate reconstruction of the spatial structure of the density field traced by the points sample. It uses the Voronoi cell volumes to obtain an estimate of the local density at each sample point, and subsequently uses the Delaunay cells to define a first-order estimate of the density field throughout the sample volume. The impressive ability of DTFE to recover the details of the *multiscale nature* and *anisotropic geometry* of the cosmic web reflects the fact that the simplicial composition of the Delaunay tessellation itself is highly sensitive to these properties. A telling visual demonstration of this is shown in the Delaunay tessellation in and around a filament in a cosmological simulation, illustrated in fig. 17 in Van de Weygaert & Schaap (2009). A careful assessment of the properties of DTFE, the Delaunay Tessellation Field Estimator, revealed that it manages to trace accurately key characteristics of the structure of the cosmic web (Van de Weygaert & Schaap 2009). To a considerable precision it follows the multiscale nature of the mass distribution sampled by a point distribution. DTFE manages to do so to the extent that it even reproduces the fractal nature of a point distribution organized in a nested sequence of ever denser point clumps. Equally important is that DTFE manages to reproduce accurately the shape of the local point distribution, and hence manages to retain the planar geometry of wall-like distributions and the elongated nature of points organized along filaments. In addition, it yields the corresponding volume-covering volume-weighted velocity field traced by the points (Bernardeau & van de Weygaert 1996; Romano-Díaz & van de Weygaert 2007).

APPENDIX B: ALPHA SHAPES

The self-adaptive nature of Delaunay tessellations with respect to local density and shape of a spatial point distribution assures that the key characteristics of the cosmic web are retained. This also concerns all information on the topology of the structure traced by the sample points. It naturally leads to the concept of the scale-based filtration of a Delaunay tessellation, defining simplicial complexes known as *alpha shapes*. These form the basis for a well-defined framework for the analysis of the multiscale geometric and topological nature of the spatial mass distribution sampled by the sample points (Edelsbrunner, Kirkpatrick & Seidel 1983; Edelsbrunner & Mücke 1994; Edelsbrunner & Harer 2010).

Alpha shapes are a well-known concept in Computational Geometry and Computational Topology (Dey, Edelsbrunner & Guha 1999;

Zomorodian & Carlsson 2005; Rote & Vegter 2006; Edelsbrunner & Harer 2010). They were introduced by Edelsbrunner and collaborators (Edelsbrunner, Kirkpatrick & Seidel 1983; Edelsbrunner & Mücke 1994). They involve a generalization of the convex hull of a point set and are concrete geometric objects that are uniquely defined for a particular point set on a scale parametrized by the scale parameter α. They have found a wide array of scientific and engineering applications, including pattern recognition, digital shape sampling and processing, and structural molecular biology (e.g. Liang et al. 1998a, b). For example, in the latter they have been used in the characterization of the topology and structure of macromolecules. They were also central in our earliest studies of Betti number characteristics and persistent topology of the cosmic mass distribution and the cosmic web (Eldering 2005; Van de Weygaert et al. 2010, 2011).

B1 Alpha shapes: definition

Alpha shapes form the core of a well-defined framework for the analysis of the multiscale geometric and topological nature of the spatial mass distribution traced by the sample points (Edelsbrunner, Kirkpatrick & Seidel 1983; Edelsbrunner & Mücke 1994; Edelsbrunner & Harer 2010). A visual impression of a two-dimensional alpha shape filtration – for a distribution of haloes – is provided by Fig. 1. Where for $\alpha = 0$ (top left) we only observe the points without any connections, this quickly changes when we increase α to 3 or 6 h^{-1}Mpc (top centre and right) and see the emergence of a connected network. The illustration also elucidates the link between alpha shapes and the homology of a point distribution. For example, we see that tunnels are formed when, at a certain value of α, an edge is added between two vertices that were already connected.

Rigorously defined in terms of the scale factors α, alpha shapes are subsets of Delaunay tessellations that represent a scale-based filtration of the distance field. The Delaunay tessellation filtration by means of alpha shapes proceeds as follows. Centred around each generating sample point, i.e. around each vertex of the Delaunay tessellation, a sphere of radius α is drawn. Subsequently we identify the simplices of the corresponding Voronoi tessellation – the dual of the Delaunay tessellation – that are located within or intersect with the *alpha* sphere union. The *alpha complex* consists of all Delaunay simplices that record (are dual to) these Voronoi simplices. The *alpha shape* for that specific value of α is the union of simplices in the alpha complex. For telling explanatory figures we refer to works by Edelsbrunner & Harer (2010) and Van de Weygaert et al. (2011).

B2 Alpha shapes: Delaunay filtrations and persistent topology

The ordered set of alpha shapes constitute a distance filtration of the Delaunay simplicial complex $\mathcal{D}$

$$\emptyset = \mathcal{D}_0 \subseteq \mathcal{D}_1 \subseteq \mathcal{D}_2 \subseteq ... \subseteq \mathcal{D}_m = \mathcal{D}. \tag{B1}$$

The alpha shapes $\mathcal{D}_k$ are hierarchical subsets of $\mathcal{D}$ that geometrically encode the change in topology throughout the simulation box, and constitute the *alpha complex*. Important is that they are homotopy equivalent to the distance field. It means that the topology of the alpha shape simplicial complex is topologically entirely equivalent to the continuous distance field (Van de Weygaert et al. 2011).

Instead of studying the topology at a single length-scale, alpha shape topology probes the topology of the distance field along its set of filtrations along the varying length-scale α. The topological information in the simplicial alpha complex is extracted by assessing the connectivity of the various simplicial components. By following

how the connections between points and simplices – within the setting of the corresponding Voronoi and Delaunay tessellation – depend on the scale parameter α we arrive at a natural description and rigorous definition of the multiscale topology. As the filtration length-scale α increases, the sample point distribution proceeds through a process in which we see a growing number of Delaunay simplices getting connected as they get embedded in the corresponding alpha complex. The increasing filtration length involves a continuously changing population of different separate simplicial complexes and the corresponding structural entities they represent.

Proceeding through the entire value range of the scale parameter α, we may follow the creation and destruction of the individual topological features. As new structures emerge, existing structures merge and other structures get annihilated, the number of (super)clusters of haloes, filaments and voids will continuously change. To obtain insight into the connection between the growing simplicial (alpha shape) complex and the topology of the spatial structure, it is illustrative to consider what happens as more points get added while the simplicial complex grows (see Van de Weygaert et al. 2011). Cycling through the simplices of the alpha complex, we encounter the following sequence. When a point is added to the alpha complex, a new component is created: the zeroth Betti number β_0 is increased by 1. In this study, the alpha complex starts out with all haloes added at $\alpha = 0$, and each component corresponds to an individual halo. When adding edges between pairs of points, the components become (super)cluster of haloes. The process of adding edges between points can have two outcomes. Either both points belong to the same, or to different components of the current complex. When they belong to the same component, the edge creates a new tunnel, increasing the one-dimensional Betti number by 1. The creation of a filament corresponds to the creation of a tunnel. When the edge connects two different components, two (super)clusters merge and β_0 decreases by 1. When a triangle gets added, it may complete the enclosure of a void or it may close a tunnel. In the first case, a new void is created and β_2 increases by 1. In the latter case, β_1 is decreased by 1 as the number of tunnels and corresponding filaments decreases. Finally, when a tetrahedron is added, a void is filled. This means that the two-dimensional Betti number β_2 is lowered by 1.

By assessing the value of α at which a feature is created, the *birth value* α_b, and the value at which a feature is destroyed, the *death value* α_d, we follow the existence of (super)cluster complexes, of filamentary features and of enclosed voids (see, e.g. Wilding et al. 2021). The difference between the creation and the destruction value,

$$\pi := \alpha_d - \alpha_b \tag{B2}$$

denotes the *persistence* π of a topological feature. By plotting the values of α_d versus α_b for all topological features present in a spatial point distribution, for each dimension $d = 0, 1, \ldots, D$, a highly detailed and idiosyncratic depiction of the topological structure is obtained. These *persistence diagrams* not only quantify the topology in terms of a few summarizing parameters, but in terms of diagrams that contain information on every individual topological feature present in the sample point distribution.

Analogously, we can compute the Betti numbers (β_0, β_1, and β_2) of the distance field as a function of the length-scale α. The resulting Betti curves (i.e. the Betti numbers for all scales) provide an overview on the complete structure of critical points, and on how the global topology differs for varying scales. While Betti curves inform us of the overall topology (since they measure the total number of topological features as a function of α), persistence diagrams allow us to precisely track at what length-scales each of these topological features is formed and for how long they *persist*.

B3 Alpha shapes illustrated

The full potential and richness of the topological information contained in alpha complexes may be appreciated from the three-dimensional situation. Figs. B1 and B2 provide illustrations of the alpha shapes in a 30 h^{-1}Mpc thick slice through the P-Millennium simulation, for three different values of α.

Fig. B1 shows the alpha shapes in slices through the *P*-Millennium simulations. They are slices with the full width of 542.16 h^{-1}Mpc of the simulation box, at three different values of the average halo–halo distance (one-third, one-half, and two-thirds). Fig. B2 provides a supplementary view of the simplicial structure, including an additional slice at $\alpha = 0.6$ (a quarter of average halo–halo distance). Along with the full slices, the right-hand column shows zoom-ins of 100 h^{-1}Mpc width. The zoom-in region is indicated in the full slices in the left-hand column.

Alpha shapes trace the intricate multiscale structure, diverse morphological features and their mutual connections of points in a web-like network. From the figures we may appreciate the natural means by which alpha shapes follow the multiscale nature and topological structure of the cosmic web (for reference also see fig. 4 in Van de Weygaert et al. 2011). The figures show the emergence of the cosmic web as traced by dark matter haloes. For small values of α (see the top row of Fig. B2) the structure consists of largely disconnected clusters of haloes and their filamentary extensions. Whereas the overall arrangement of clusters amidst filamentary and wall-like extensions is clear, these 'supercluster' complexes still appear to be mostly singular, isolated features. They are separated by large underdense regions, in which tenuous largely elongated features are visible. These are the denser parts of large-scale filamentary bridges and tendrils. Only at higher values of α we see that the high-density supercluster islands start to get connected via filamentary bridges, connections that become stronger, more elaborate and more structured as more and more features get connected in an emerging pervasive network. It is in particular the central panel of Fig. B1 that offers the most detailed view of the intricacies of the fully connected spatial network. At even higher values of α the lower density areas start to fill in the cavities in the spatial network. This yields the reverse impression of the structure at lower α values: it is the underdense void regions that turn into ever more isolated and encapsulated cavities. The third panel of Fig. B1 depicts this phase of the alpha complex.

The zoom-in panels of Fig. B2 are centred around a low-density void region. It provides a telling illustration of the substructure of voids and of the signature of the corresponding multiscale topology in the alpha complex. An almost empty region at the lowest α value, gets gradually structured towards higher α values. At first this concerns isolated clumps, which towards the higher α values get connected in elongated features (third panel). Ultimately, towards the highest α values, these features enclose small subvoids of the initially large void (fourth panel). In short, the four zoom-in panels of Fig. B2 provide visually compelling evidence for the ability of alpha shapes to probe the detailed and intricate multiscale structure of the cosmic web and its components.

Translated into the corresponding Betti number signature, the first panel of Fig. B1 is marked by the dominance of β_0, while the central panel corresponds to high values of β_1 and a diminished β_0 signature due to the merging of supercluster islands into the cosmic web. The third panel has a dominant β_2 signature, as a result of the increasing abundance of fully enclosed void cavities, while the emergence of their wall-like boundaries implies a substantial decrease of β_1 due to the disappearance of filament-related tunnels in the corresponding alpha shapes.
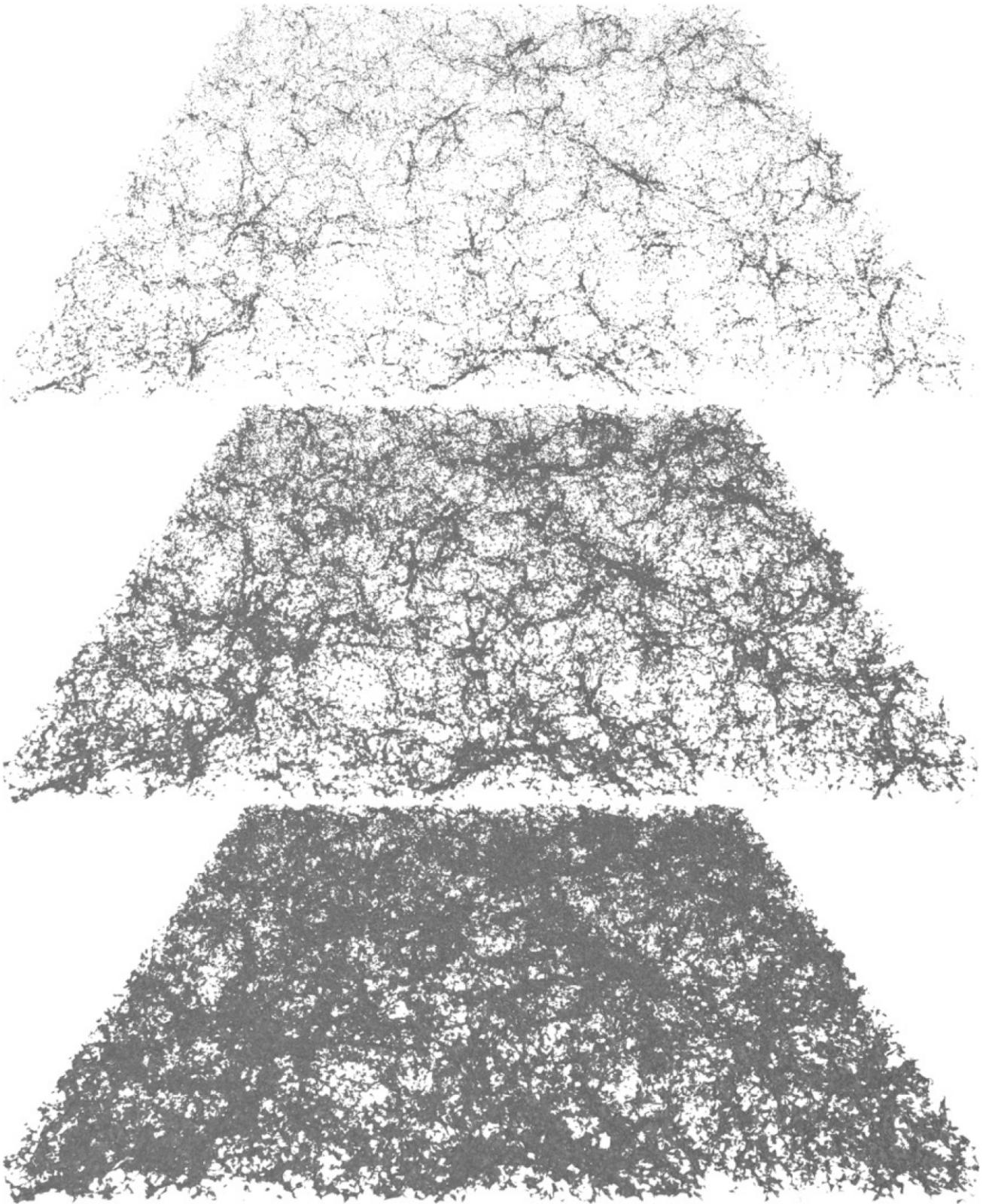
Figure B1. Three-dimensional alpha shapes. From top to bottom, we show the three-dimensional alpha shapes of a slice of the full halo distribution for $\alpha \approx$ 0.8, 1.2, and 1.6 h^{-1}Mpc (corresponding to one-third, one-half, and two-thirds of the average halo–halo distance).
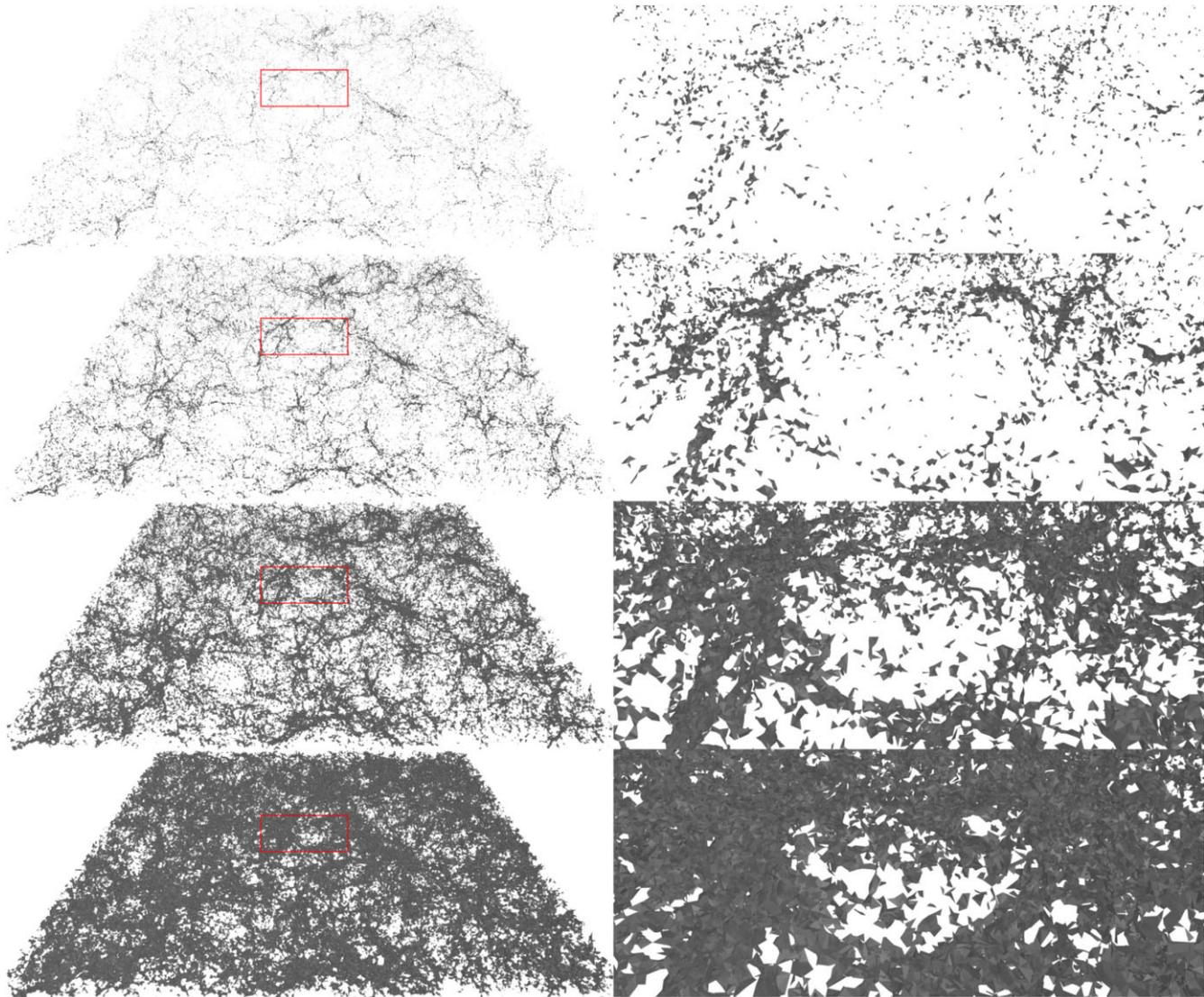
Figure B2. Three-dimensional alpha shapes. From top to bottom, we show the three-dimensional alpha shapes of a slice of the full halo distribution for $\alpha \approx 0.6$, 0.8, 1.2, and 1.6 h^{-1}Mpc (with the additional top slice corresponding to a quarter of the average halo–halo distance). The right-hand column shows a zoom-in of roughly 100 h^{-1}Mpc width, which is marked by a red rectangle in the corresponding left-hand panel.

APPENDIX C: DISTANCE FIELD ANALYSIS – ALPHA SHAPES VERSUS GRAPH METHODS

The sampling of the distance field is a well-known and suggestive method for quantifying the cosmic matter distribution probed by the discrete distribution of haloes or galaxies. It is familiar in the context of both halo and clump classification, as well as in a range of techniques for assessing aspects of the spatial distribution of galaxies. The FRIENDS-OF-FRIENDS algorithm (Davis et al. 1985; White, Cohn & Smit 2010), and elaborations thereof (Behroozi, Wechsler & Wu 2013), is a widely applied tool for the analysis of *N*-body simulations and identification of haloes.

One of the earliest examples of its exploitation in the context of the large-scale structure of the Universe is that of the percolation analysis by Zeldovich, Einasto & Shandarin (1982), seeking to establish aspects such as at which scale connected structures form that span the entire volume, or at which scale all points in a sample get connected. A more intricately and uniquely defined structure is that of the Minimum Spanning Tree (MST), a construct stemming from graph theory. It links *all* points in a sample by means of a *graph* with a minimal total length. It has now a range of applications in the analysis of the cosmic galaxy distribution. Bhavsar and collaborators (Barrow, Bhavsar & Sonoda 1985; Graham, Clowes & Campusano 1995) introduced MSTs to quantify the filamentary nature of the galaxy distribution. Based on MSTs, Colberg (2007) developed a formalism to identify and classify the web-like distribution of haloes and galaxies, an approach that got further elaborated on by Alpaslan et al. (2014b). Their MST-based formalism has been applied for the identification of filaments and void regions in the GAMA survey (Alpaslan et al. 2014a).

Another approach to the use of MSTs concerns an assessment of a range of statistical measures of the MST edge length distribution. Naidoo et al. (2020) used this in an attempt to quantify higher order clustering properties. Bonnaire et al. (2021) also used aspects of graph theory, and a keenly defined statistical procedure to regularize the MST, to obtain a technique that identifies in particular the dominant filaments in two- and three-dimensional galaxy distributions.

In the more general setting of *graph theory*, a range of studies have been using the statistics of edges connecting the points in the spatial particle, halo, or galaxy distribution to extract clustering measures. Statistical moments of the edge length distribution are proposed to quantify aspects of the higher order spatial clustering of the sample points. Expecting that these measures pertain to aspects of their web-like distribution, *network analysis* is obtaining increasing attention (Hong & Dey 2015; Hong et al. 2016; De Regt et al. 2018; Tsizh et al. 2020) in large-scale structure studies.

These applications underline the substantial amount of information contained in the distance field, sampled by the mutual distances between the sample points. Nonetheless, it is through the unique, transparent, geometrically and topologically ordered means by which *alpha shapes* are processing the vast information content of the sample point distribution, and the connections between the points, that the measures extracted can be related to visually accessible and topologically and geometrically meaningful information. They do so in an entirely natural fashion, without the need for ad hoc parameters or fitting to heuristic mathematical structures.

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