Draft version January 31, 2020
Typeset using LaTeX **preprint2** style in AASTeX62

The cosmic web through the lens of graph entropy

María Valentina García-Alvarado,[1] Jaime E. Forero-Romero,[1] and Xiao-Dong Li[2]

[1]*Departamento de Física, Universidad de los Andes, Cra. 1 No. 18A-10 Edificio Ip, CP 111711, Bogotá, Colombia*
[2]*School of Physics and Astronomy, Sun Yat-Sen University, Guangzhou 510297, P. R. China*

(Accepted January 31, 2020)

Submitted to ApJ

ABSTRACT

We present the graph entropy of the cosmic web as a new scalar to quantify the large-scale structure of the Universe. We define the entropy, S, from the probabilities, p_n, of a node having n connections, $S = \sum -p_n \log_2 p_n$. The basis of our quantitative analysis is the β-skeleton graph constructed on dark matter halos from cosmological N-body simulations. We find that the graph entropy varies between 1.6 and 3.2 Shannon. The minimum entropy is achieved by the 2-skeleton, also known as the Relative Neighborhood Graph. We also explore the influence of survey geometry, cosmic variance, redshift space distortions, redshift evolution, cosmological parameters, clustering and spatial number density on the graph entropy. Cosmic variance shows the least important influence on the entropy with effects on the order of 10^{-3} Shannon. The survey geometry, redshift space distortion, cosmological parameters, redshift evolution and clustering vary the entropy on the order of 1×10^{-2} Shannon. The strongest influence on the graph entropy is presented by random points versus clustered halos at different number densities with changes up to 0.2 Shannon.

1. INTRODUCTION

The cosmic web is the outstanding pattern on the spatial galaxy distribution on its largest scales.

Graphs have been an alternative to describe the cosmic web. (voronoi, beta-skeleton, void finders, other graphs).

However, it is difficult to summarize the complexity of these graphs into scalar quantities to manipulate. Some options have been the average connection length, the average direction of the connections (relevant in the case of redshift space distortions) or the average number of connections.

In this work we use the

The configuration entropy has been recently used (Pandey & Das 2019)

2. GRAPH ENTROPY

2.1. *The Beta-Skeleton Graph*

The β-Skeleton is a graph that determines the connectivity between pairs of points as a function of the real parameter β. It establishes that only if there is no other node in the space between two fixed points, these two nodes are related. As seen in figure 1 the dimension of the void between two nodes increases with the parameter β. For the smallest value ($\beta = 0$) the region without any other node is just a line that joins both points. As the parameter gets larger the restricted zone grows until it gets

large enough so no pair of points are related. In particular, when $\beta = 1$, the β-Skeleton results with the same structure as the Gabriel Graph in which all the points are related with each other, in this case a complete graph is constructed.

As the required void gets larger, it gets harder to find pairs of points isolated enough to satisfy the condition and edges are lost with respect to a β-Skeleton graph constructed with a smaller value of β. To illustrate this, Figure 2 shows the constructed structure for three different values of β.

2.2. *Entropy Definition*

Now that the structure of the β-Skeleton graph has been shown,the resemblance with the cosmic web is clear as both seem to follow a filamentary pattern. For this reason, we will analyze data from cosmological simulations to measure and constrain the characteristics that describe the large scale structure of the universe, such as the behaviour of the cosmological parameters, the effects of redshift space distorsion (RSD), the changes found between different redshift, among others.

3. METHODS

3.1. *Simulations*

Abacus Cosmos (Garrison et al. 2018) is a project that contains N-body simulations of dark matter halos. In particular it includes simulations with the standard and different cosmologies.

We selected the *AbacusCosmos_720box* as the source of data for the probes with the standard cosmology. This simulation set is a box with size of 720 Mpc/h and is divided in 20 subsets to cover all the space.

We also considered the 720 Mpc/h sized box, *AbacusCosmos_720box_planck*. This simulation set is divided in 40 files, each of them represent a different cosmology which means that it was developed under certain combinations of cosmological parameters.

Between all the information given by the simulations we just took into account, for the selection of our points, the positions and masses of the dark matter halos.

3.2. *Mock Catalogs*

As it was mentioned above, all of the data obtained from Abacus Cosmos was arranged in boxes. From these we selected the more massive dark matter halos with a maximum circular velocity $\geq$300 km s^{-1} . From the resulting points, we discarded some points to get only the halos that formed a spherical geometry(sphere) in the center of each of the files from the two simulation sets. All of the spheres have the same dimensions: A radius of 300 Mpc h^{-1}.

We used additional sets of data with similar characteristics as the spheres extracted from Abacus for most of the probes that we will present next. We prepared points that considered the effects of geometry of the set of points, the difference with random generated points and the redshift distorsion effect. In particular, for geometry effects we chose from each sphere with standard cosmology the points corresponding to a shell with radius of 300 Mpc. Both geometries are shown in Figure **??** (tocaria cambiarlo por los radios que si son). We note that even though the data shown in the figures is presented in two dimensions, all the following analysis is done in three dimensions.

3.3. *Procedure*

We applied the β-Skeleton algorithm[ref] to find all the pairs of dark matter halos that are related according to a specific value of the parameter. For this selection, it takes into account the geometric restriction in three dimensions(in a similar way as seen in Figure 1, but taken into three dimensions). The graph was constructed with the values of $\beta = 1,\ 1.5,\ 2,\ 2.5,\ 3,\ 3.5,\ 4$

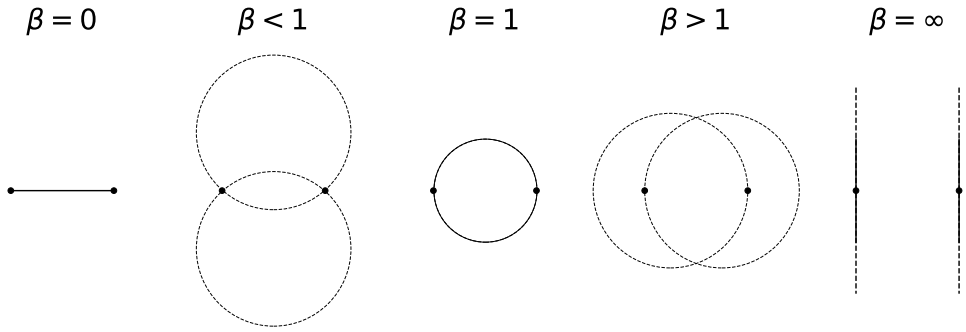

Figure 1. Exclusion regions between two points in the β-Skeleton definition for different β values. The exclusion region is defined by the intersection between the two circles of the same diameter that pass by the two points, its diameter is $D = d/\beta$ for $\beta \leq 1$ and $D = \beta d$ for $\beta > 1$, where d is the distance between the two points. If this exclusion region is empty from any other third point, then the two points are linked. For $\beta = 0$ the exclusion region is a line connecting the dots, while for $\beta = \infty$ the exclusion region becomes the semi-infinite region delimited by two parallel lines. The case of $\beta = 1$ corresponds to the Gabriel Graph.

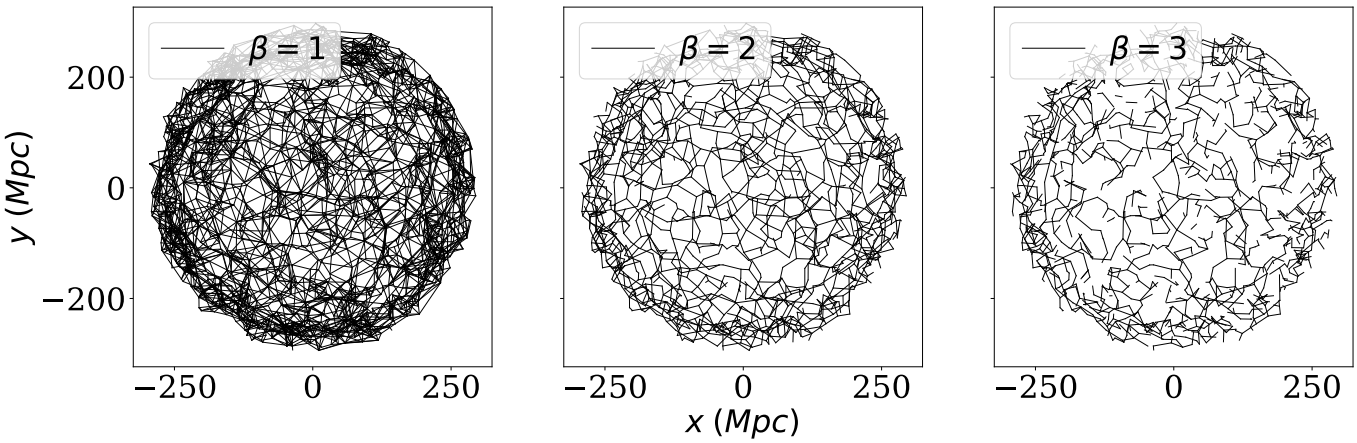

Figure 2. β-Skeleton graph for a set of dark matter halos arranged over a shell with inner radius 250 $\mathrm{Mpc}\,h^{-1}$and outer radius 300 $\mathrm{Mpc}\,h^{-1}$. Each panel correspond to different values values of β (1, 2 and $\beta = 3$). The case of $\beta = 2$ corresponds to the Relative Neighbor Graph. As β increases links between nodes are lost. The probability of being a node with n connections changes with β.

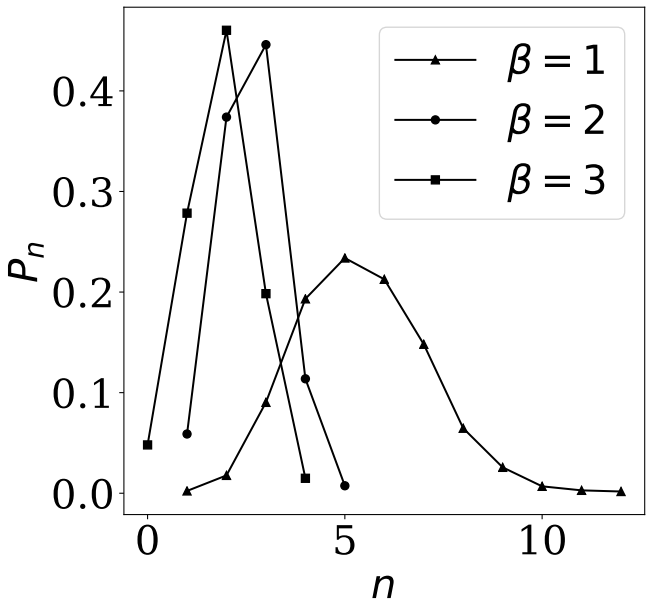

Figure 3. Probabilities of having n connections, p_n for three different values of β. The input dataset corresponds to the shell in Figure 2. We use the entropy to summarize the changes in the p_n distribution as a function of β.

and 5 for every data set we considered and was already mentioned.

For each data set and each *beta*, we computed the grade of connectivity (the number of edges) of all the nodes that composed it. With these we got the probabilities of having a determined number n of edges. We estimate these probabilities following the expression $P_n = \frac{N_e}{T_n}$, where N_e is the number of nodes with $n \geq 0$ links and T_n is the total number of nodes in the data set. An example of the probabilities (P_n) found for one of the data sets and for three values of *beta*, is shown in Figure 3.

As it was described before, for convenience, we will use the graph entropy as the base of the further analysis. Hence, we use the probabilities already found in the calculation of this quantity which is given by Eq.1

$$S = \sum_{P_n > 0} -P_n \log_2 P_n \qquad (1)$$

The product of the calculation of the entropy on all the data sets prepared, allowed us to describe the effects of the variation of the parameter *beta* on various characteristics of the dark matter simulations. The results that will be presented are based on the study of the effects of Cosmic Variance, RSD, computation of random points entropy, redshift evolution, and variation of: Geometry, Cosmological Parameters and data clustering degree.

4. RESULTS

Before getting into showing the results, it is important to describe the entropy graph in which all the conclusions will be based. A first sight of this can be seen in the left plot of the Figure 4 were the entropy, as a function of β, follows a decreasing pattern that seems to be interrupted when $\beta = 2$, which represents a clear local minimum of the curve. This remarkable feature differentiates the point at which the constructed graph goes from being connected, to have completely isolated nodes (that are unreachable from any other node).

4.1. *Cosmic Variance*

We described the effects of cosmic variance by comparing the entropy of 20 spheres located at different regions of the simulated space with the standard cosmology, without considering the effects of RSD and at $z = 0.1$. The entropy graph that illustrates this corresponds to Fig.4, in which the entropy curve of all spheres overlap, and the relative entropy of each sphere with respect to a sphere of reference has a magnitude of 10^{-3} Shannon. This quantity is small compared to the entropy values. Hence, it tells us that the entropy is weakly dependent on the variance of the points under consideration.

4.2. *Geometry*

We evaluated the effects produced by the geometry of the points on the entropy. For this, we computed the entropy for 20 shells located at

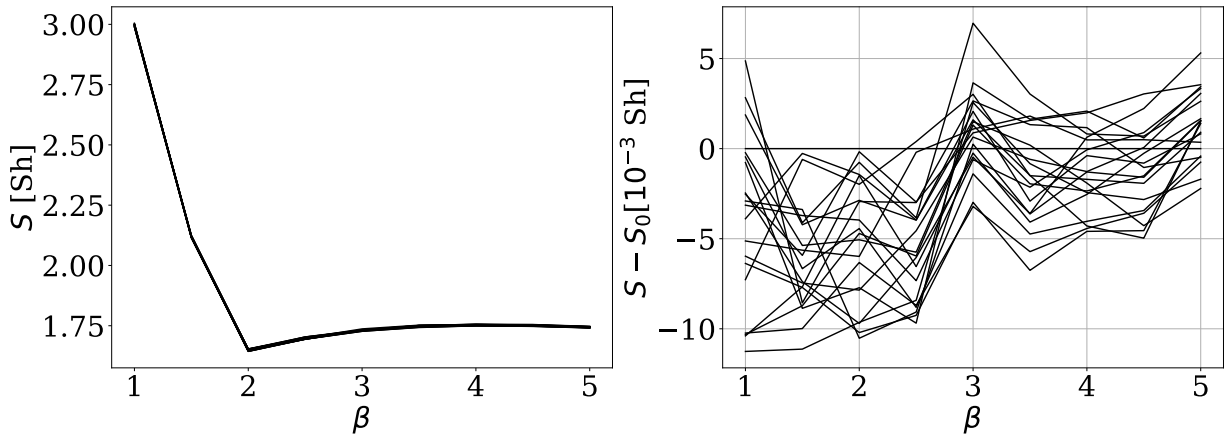

Figure 4. Influence of the cosmic variance on the entropy. This calculation is performed on 20 different spheres at $z = 0.1$. The left panel shows the absolute entropy values, while the right panel shows the entropy relative to an arbitrary reference simulation.

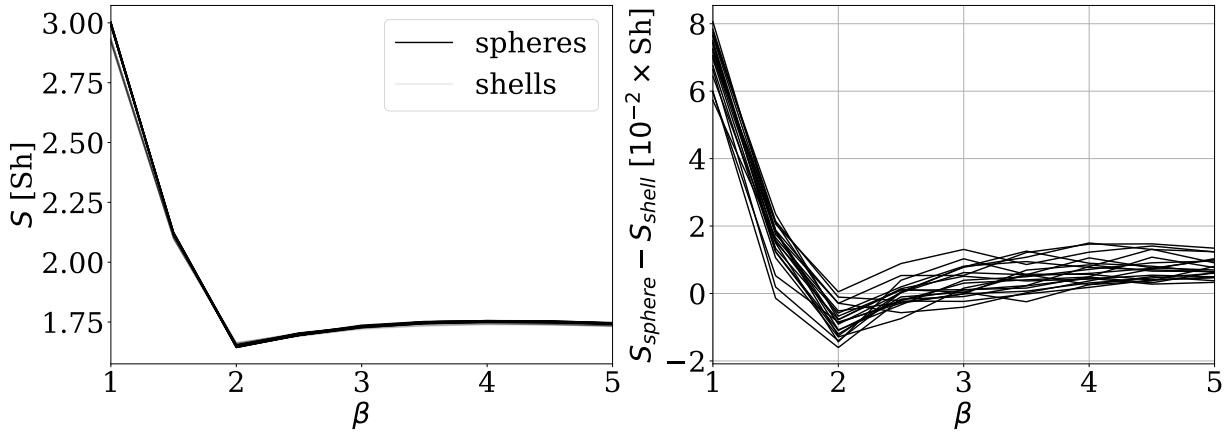

Figure 5. Influence of the survey geometry on the entropy. This calculation is performed on 20 different spheres and shells at $z = 0.1$. The left panel shows the absolute entropy values, while the right panel shows the change between the corresponding sphere and shell.

different regions of the simulated space with the standard cosmology, without considering the effects of RSD and at $z = 0.1$. This entropy graph and the obtained for 20 spheres were compared, as it is shown in Fig.5. Again, the entropy curve of spheres and shells seem to be overlaped, and the local minimum at $\beta = 2$ is conserved. The relative entropy of the shells with respect to the corresponding concentric sphere has a magnitude of 10^{-2} Shannon. This value is small compared with the actual entropy, which means that, as seen with the cosmic variance, the geometry variation has a weak effect on entropy.

4.3. *Redshift space distortion*

We measured the influence of redshift space distorsion (RSD) by observing the variation between the entropy graph of two sets of 20 spheres, both developed with the standard cosmology and at $z = 0.1$, but one of them considers RSD. Figure 6 exposes what was found: There is no clear distinction between both curves as they seem to follow exactly the same behaviour as β increases. The relative entropy of the spheres with RSD with respect to the corresponding concentric sphere without RSD is in the order of 10^{-2} Shannon. The changes induced by RSD are not significant compared to the absolute entropy values. We can tell that entropy is also weekly dependent on Redshift Space Distorsion.

4.4. *Redshift evolution*

To constrain the influence of redshift evolution on the entropy, we analyzed six spheres with the same spacial coordinates, without considering RSD, with the standard cosmology, but at different redshift $z = 01, 0.3, 0.5, 0.7, 1, 1.5$. We computed the β-Skeleton for $\beta = 1, 2, 3$ and its corresponding graph entropy. Fig.7 illustrates the results obtained. Each β shows a different behaviour of the entropy as a function of the redshift. The changes of the entropy with respect to $z = 0.1$ are in the order of 10^{-2} Shannon.

4.5. *Cosmological Parameters*

We measured the influence of the variation of the cosmological parameters $H_0, \Lambda, \Omega_m, n_s, \sigma_8, \omega_0$ on the entropy. We took into account for this constraining 40 spheres, each with different combinations of the considered parameters, at $z = 0.1$ and without RSD effects. In Fig.8 we present the product of the entropy computed from the $\beta = 1$ graph as a function of every cosmological parameter. It is evident that the different parameters induce changes on the entropy of magnitude 10^{-2} Shannon. Some of the parameters expose a clear correlation with the entropy. It is seen in the ρ value which is the Spearman's rank correlation coefficient. The values $\rho = 1, -1$ indicate a direct and inverse correlation, respectively. Hence, the strongest correlation is observed with σ_8 and ω_0. The entropy can be weakly or strongly influenced by the variation of the cosmological parameters, depending on the parameter taken under consideration.

4.6. *Clustering degree*

We use five different spheres at $z = 0.1$ with different minimal cuts on the maximum circular velocity to produce samples with different clustering properties. The cuts are at 300, 350, 400, 450 and 500 $\mathrm{km\,s^{-1}}$, which translate correlation lengths of 6.3, 7.3, 7.8, 8.4 and 9.7 $\mathrm{Mpc}\,h^{-1}$, respectively.

Figure 9 shows the the influence of increasing clustering degree is on the order of 10^{-2} Sh. Although this values is relatively low, for the $\beta = 1$ we find a clear negative correlation between the entropy change and the correlation length. As the correlation length increases, the entropy decreases. This correlation is singled out in Figure 10

As an extreme case of clustering difference, we measured the entropy in 20 spheres with a

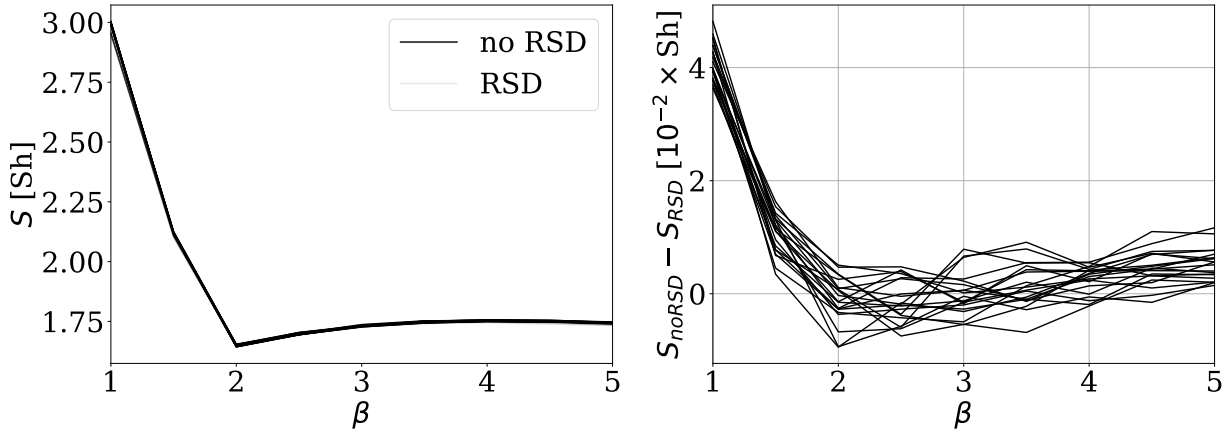

Figure 6. Influence of the Redshift Space Distortions (RSD) on the entropy. This calculation is performed on 20 different spheres at $z = 0.1$. The left panel shows the absolute entropy values, while the right panel shows the change between the corresponding sphere with and without RSD effects. The changes induced by RSD are on the order of 10^{-2} Shannon.

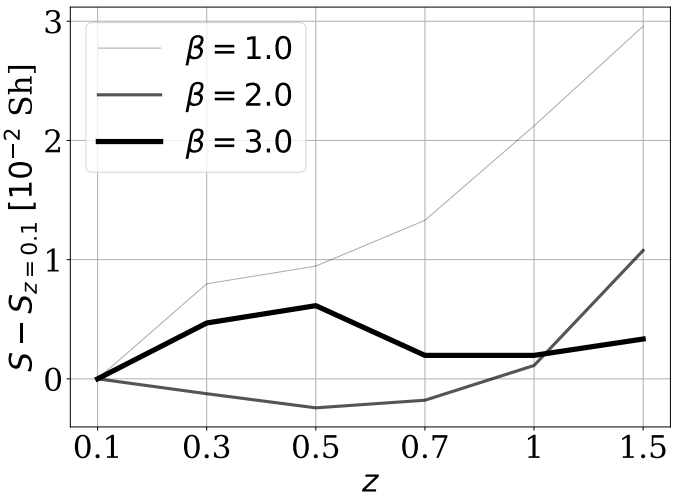

Figure 7. Redshift evolution of the entropy for different values of β. This calculation is performed on a single spheres at different redshifts. Different values of β show different redshift dependencies. The strongest one is found for $\beta = 2$. Overall the entropy changes on the order of 10^{-2} Shannon as a function of redshift.

random distribution of points. Each one of the spheres has the same number of points as each one of the spheres with halos at $z = 0.1$ with a maximum circular velocity cut at 300 km s^{-1}.

Figure 11 shows the clear difference between random The difference between entropy in random points and entropy resulting clustered data reaches up to 0.2 Sh.

4.7. *Number densities*

We prepared and additional dataset where we randomly sample a percentage (between 10% and 90%) of the points from the original random and clustered spheres. Although the number density changes, the different points have the same correlation function as they are dominated by dark matter halos with circular velocities ≈ 300 km s^{-1}.

However, the entropy changes are large. Figure 13 shows the entropy as a function of β for the spheres with different number densities. The changes in entropy between spheres with the percentages of points are in the range of 0.05 to 0.20 Sh.

These changes are not entirely explained by the different number densities. To test this we computed the entropy on the spheres with random distribution of points with the exact same number density. The results are summarize in Figure **??** In this case the changes on the entropy between the curves have a magnitude of a few 10^{-2} Sh, always smaller than the changes in the clustered points.

To summarize, the entropy in this case is sensitive to changes in the dataset that are invisible to the two-point correlation function. Furthermore, these changes are only visible on points that are not randomly distributed. One of the properties that changes with the number density is the visual appearance of the cosmic web on the dataset. As the number density decreases the voids in the distribution become larger and the filamentary patterns thinner.

In order to quantify this statement we measured the typical void size in the different datasets. We computed the void probability funcion (vpf), that is the probability of having an empty sphere with a given radius, both on the clustered and random points. Voids with a radius smaller than the mean distance between points have a probability to occurr close to unity. Larger voids on the other hand have a close probability to appear. We find the typical void size by taking a look at the ratio between the vpf on the clustered data to the vpf on the random data. This ratio peaks at the typical void size in the dataset.

The left panel on Figure 14 shows the typical size of the voids as a function of the sampling percentage of the clustered points. There is a clear correlation showing that a denser dataset presents smaller voids. On the right panel of the same Figure we show the void size as a function of the entropy difference between random and clustered points. There is a clear negative cor-

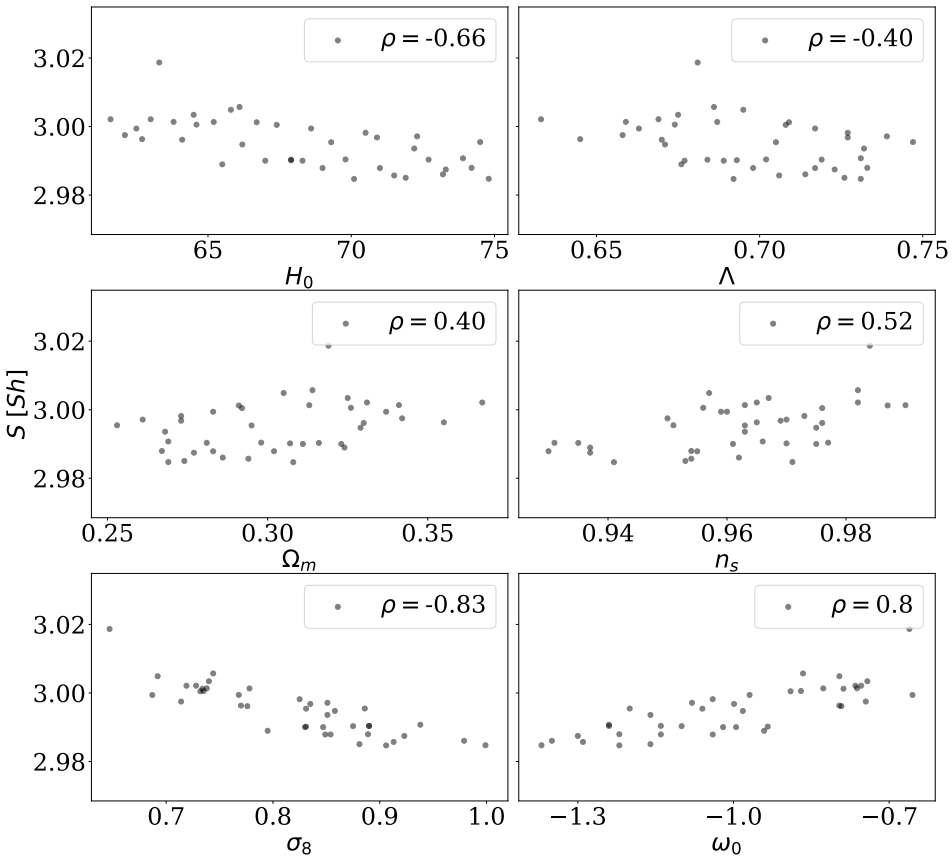

Figure 8. Influence of the cosmological parameter on the entropy. This calculation is performed on 40 different spheres at $z = 0.1$. Each panel shows the entropy as a function of a different cosmological parameter. The ρ value in the label is the Spearman's rank correlation coefficient. Different cosmological parameters induce changes on the order of 10^{-2} Shannon.

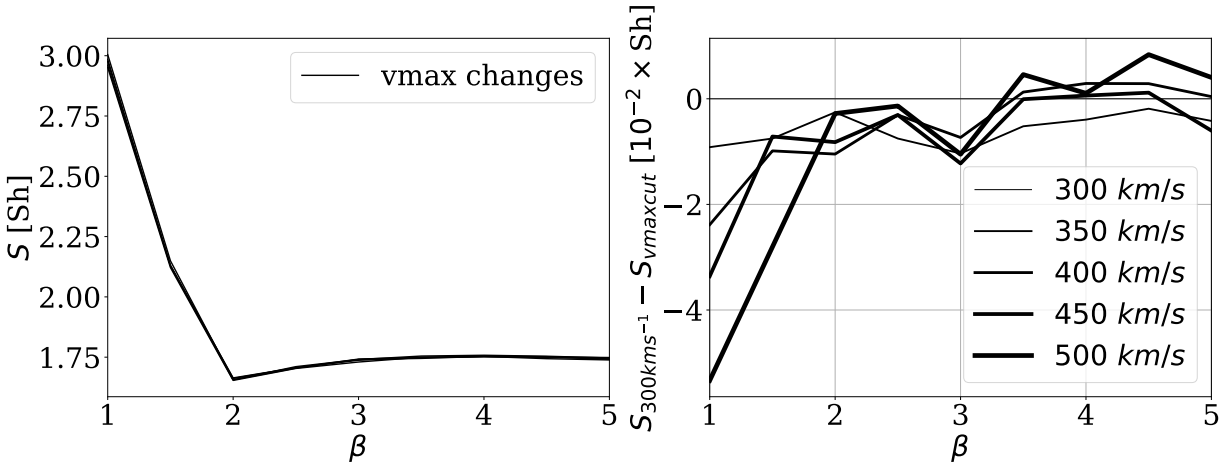

Figure 9. Influence of the clustering degree on entropy. This calculation is performed on 5 different spheres with halos more massive than five different cuts in the maximum circular velocity.

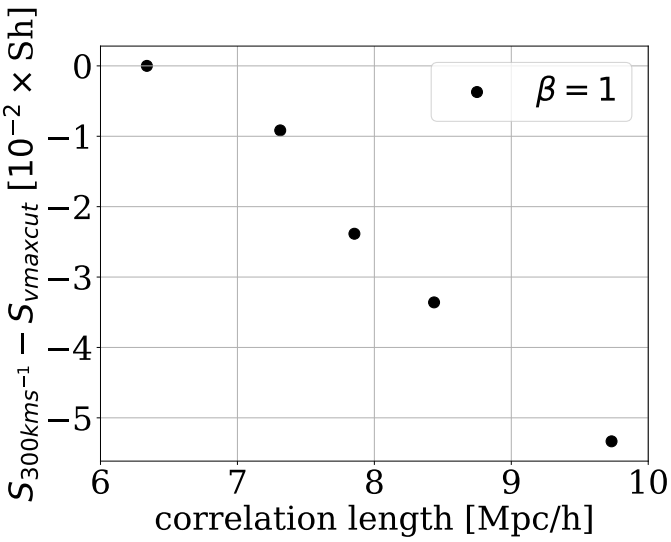

Figure 10. Correlation between entropy change and correlation length for the samples presented in Figure 9.

relation that quantifies the strong link between the entropy and the cosmic web features.

5. CONCLUSIONS

In this work we presented the entropy of a graph as a new scalar to quantify the large scale structure of the Universe. The entropy we use corresponds to the definition commonly used in information theory. It uses the probability of having n connections, p_n to build the scalar $S = \sum -p_n \log_2 p_n$. We based our analysis on the β-skeleton graph built on different mock catalogs constructed from cosmological N-body simulations.

After measuring that the entropy is of order unity and ranges between 1.6 and 3.2 Shannon, we tested the effect the influence of six major factors on the entropy. These factors were: cosmic variance, survey geometry, redshift space distortions, redshift evolution, cosmological parameters and clustering.

We found that the weakest influence is presented by cosmic variance with changes on the order of 10^{-3} Shannon. The next factors in order of importance were the survey geometry, RDS, redshift evolution, cosmological parameters and different clustering strenght with entropy changes on the order of 10^{-2} Shannon. The strongest influence on the entropy appears between clustered and random data and at different number densities. The entropy difference between random points and clustered points with different number densities is in the range of 0.05 to 0.20 Shannon. We showed that a simple topological property (the typical void size) correlated with the entropy. This shows that the graph entropy can be a robust scalar to quantify the cosmic web in simulations and observations.

Future work includes using the entropy to describe the cosmic web as mapped from current and future large scale galaxy surveys such as the Sloan Digital Sky Survey and the Dark Energy Spectroscopic Instrument (DESI). More work is needed to address different issues on other observtional effects such as survey completeness and light-cone effects. Furthermore, this definition of the graph entropy is not restricted to the β-skeleton. It is of interest to explore the results of this definition on different kinds of graphs.

REFERENCES

Garrison, L. H., Eisenstein, D. J., Ferrer, D., et al. 2018, ApJS, 236, 43, doi: 10.3847/1538-4365/aabfd3

Pandey, B., & Das, B. 2019, MNRAS, 485, L43, doi: 10.1093/mnrasl/slz029

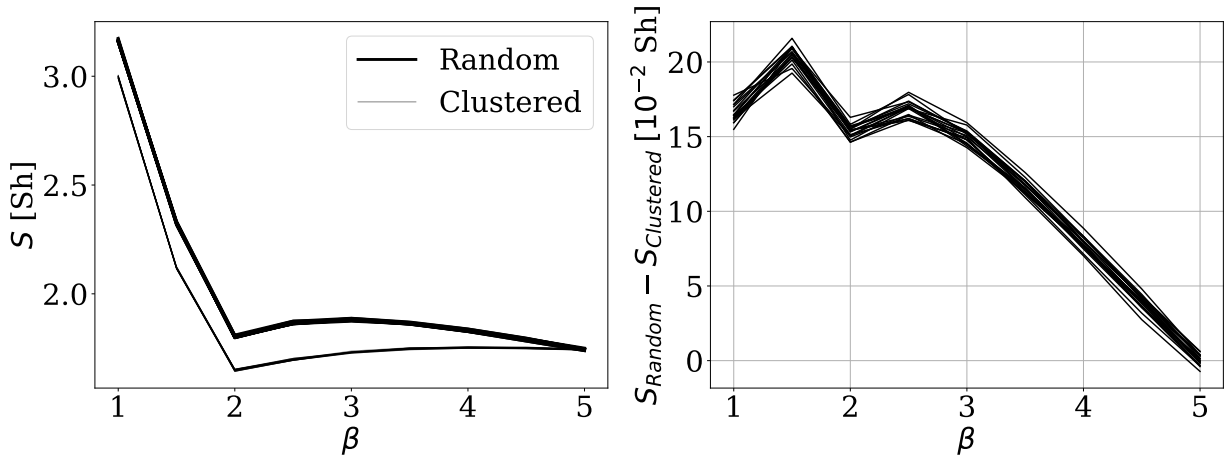

Figure 11. Influence of the clustering degree on the entropy. This calculation is performed on 20 different spheres at $z = 0.1$. Each sphere from the simulation has its corresponding random catalog with the same geometry and the same number of points. The left panel shows the absolute entropy values, while the right panel shows the change between the corresponding sphere with its random catalog. The changes between clustered and random data shows the strongest effect with differences up to 0.2 Shannon.

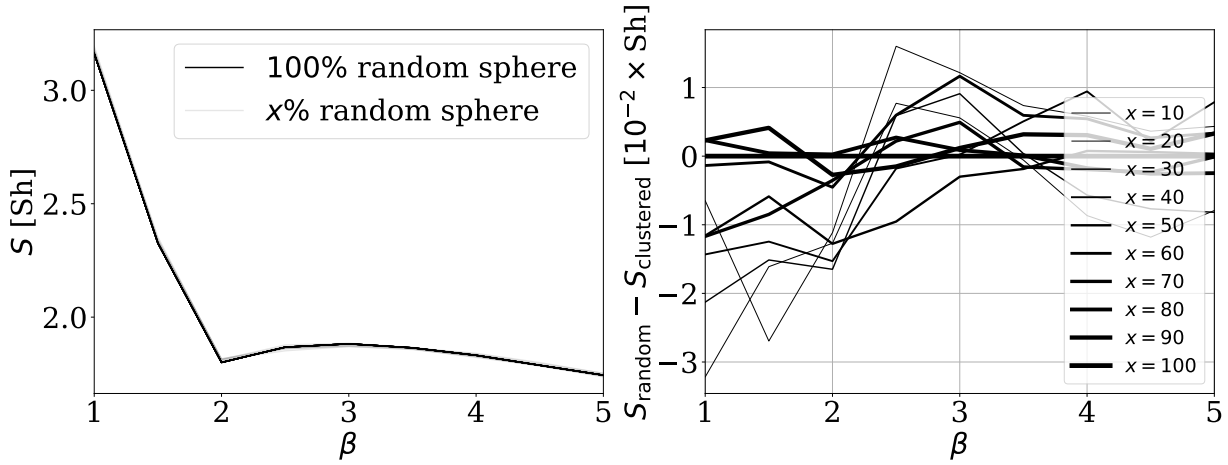

Figure 12. Influence of the number density of points on the entropy. This calculation is performed on 20 different random points distributed as spheres. A percentage, $10 < x < 100$, of these points are subsampled. The left panel shows the absolute entropy values, while the right panel shows the change between the entropy in the full dataset and the corresponding sampled catalog. The influence of the number density is on the order of 10^{-2} Shannon.

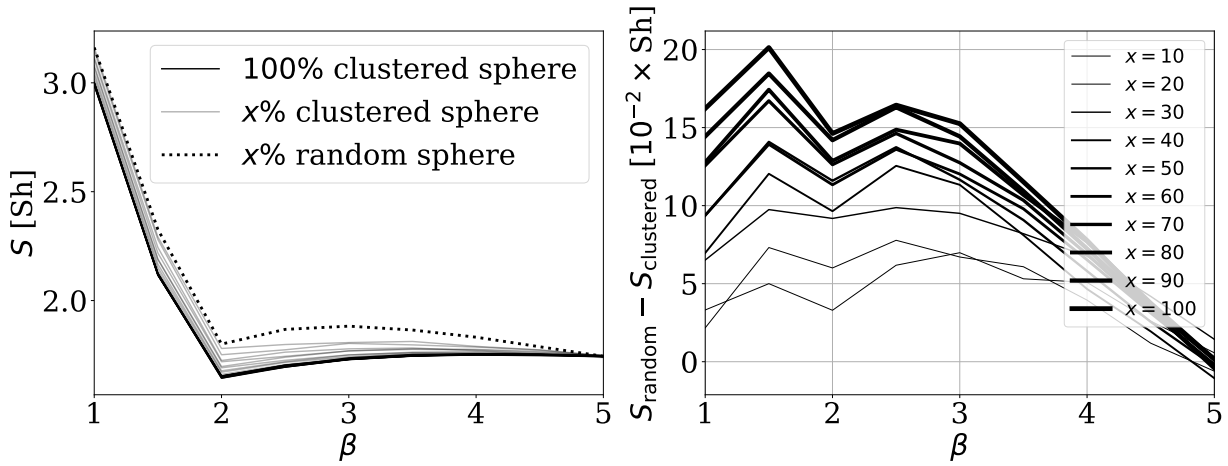

Figure 13. Influence of the number density of clustered points on the entropy. This calculation is performed on 20 different clustered points distributed as spheres at $z = 0.1$. A percentage, $10 < x < 100$, of these points are sampled. The left panel shows the absolute entropy values, there we show for comparison the results from the sampled random points in Figure 12. The right panel shows the difference between the entropy in the sampled clustered sphere and its corresponding random dataset. The changes in this case range between 0.02 and 0.2 Shannon. Different percentage of points are effectively showing a different clustering/topology in the data that is revealed by the comparison against the entropy random dataset.

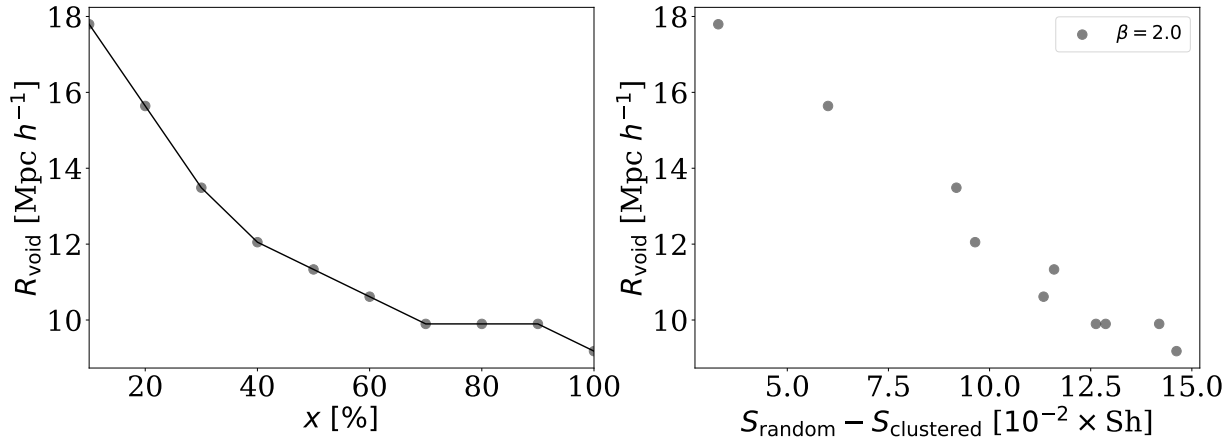

Figure 14. Typical size of the voids $R_{\rm void}$ in the clustered points as a function of the sampling percentage (left panel) and the entropy difference between the random and clustered points (right panel). This is computed with the same dataset used in Figure 13. This shows how the entropy correlates with a characteristic topological property of the dataset, in this case the typical void size.