The Cosmic Web Network

Italo Perrucci
0001087782
italo.perrucci@studio.unibo.it

Beatrice Magni
0001103483
beatrice.magni5@studio.unibo.it

September 2023*

Abstract

The Cosmic Web, the incredible structure underlying the Universe, is here analyzed from the point of view of the Complex Network methodology. First of all we build the network structure following three different methods and analyze their differences and similarities based on their network characteristics: average clustering coefficient, largest strongly connected component and small-worldness. Further on, we proceed to identify the structural components of the Cosmic Web via the widely used Voronoi tessellation method and compare it to the degree centrality method, using the first to maximize the efficiency of the second, allowing us to give some constraints for future analyses based also on different datasets.

1 Introduction

The past few decades have revealed that on scales of a few up to more than a hundred Megaparsec, the matter content of the Universe conglomerates into an intriguing cellular or weblike pattern that pervade the observable cosmos (Fig. 1). The key structural components of the cosmic mass distribution are[1]:

- clusters;
- filaments ;
- walls or sheets;
- voids;

which are not merely randomly and independently scattered features. On the contrary, they have arranged themselves in a seemingly highly organized and structured fashion, the *Cosmic Foam* or *Cosmic Web*.

Towards the end of the seventies a set of new observations started to unveil the existence of coherent structures larger than that of clusters of galaxies. Thus the supercluster paradigm [2] established itself as the new view of the large-scale distribution of matter and galaxies in the Universe. It had gradually emerged as a result of various early galaxy redshift surveys of nearby regions in the Universe (e.g. [3]) and put on a firm with the completion of the first systematic redshift survey, the CfA1 survey [4]. Along with these efforts came the unexpected finding of the first example of large cosmic voids, the Bootes void.

The importance of the analysis of the Cosmic Web resides on the information that it can give about the underlying cosmological model and about the evolution and characteristics of galaxies, giving an insight on a wild variety of scales. Also its evolution, structure and dynamics are to a large extent dependent on the nature of dark matter and dark energy. In fact, the evolution is directly dependent on the rules of gravity and therefore each of the relevant cosmological variables will leave its imprint on the topology of the Cosmic Web and on its structural elements, that is filaments, walls, clusters and voids. In particular void regions of the Cosmic Web offer one of the cleanest probes and measures of dark energy as well as tests of

*`https://github.com/itperr/Cosmic_web.git`

Figure 1: Graphic representation of the Cosmic Web, from "The Network Behind the Cosmic Web", by B. C. Coutinho et al.[5]

gravity and General Relativity.

2 Models of Cosmic Web

In the history of the Cosmic Web many algorithms have been implemented to harness the distribution of galaxies and dark matter haloes into clusters, voids, walls and filaments (for a review see [6], [7]). In this report we use the framework of *Network Analysis* to identify the constituents of the Cosmic Web and to gain insights on some of its properties. We use three different models to build the network:

- M1, the simplest one, connects two nodes with an undirected edge if the distance between them is smaller than a threshold length l;
- M2 creates a directed edge from a node to its k closest neighbors, therefore its out-degree is fixed while its in-degree may vary;
- M3 creates a directed edge from a node to the k nodes exerting the strongest gravitational pull on it in terms of the classical Newtonian law: $F = G_N \frac{M_1 M_2}{r^2}$.

Each of these models depends on a parameter, whether it is the distance threshold l for M1 or the number k of nodes to connect to for M2 and M3. We choose to set $l = 8$ for M1 and $k = 5$ for both M2 and M3, the justification of these values will be discussed in section 4.1. The resulting networks are represented in Fig. 2, where the size of each node is proportional to its degree for the undirected graph M1 and to its in-degree for the directed graphs M2, M3.

3 Data

The nodes of our network are taken from `[http://dx.doi.org/10.17876/data/2017_1]`. The dataset is a collection of more than 280k dark matter haloes identified by applying a standard FOF algorithm to the $z = 0$ snapshot of the Gadget-2 [8] dark matter only N-body simulation, with a box size of $200h^{-1}Mpc$ and 512^3 particles, with $h = 0.68$ [9]. However, since the dimension of the dataset is large and the computational demand is high, we limit ourselves to the study of a sample of the 30k nodes with largest mass, which of course results in a mass-biased sample.

4 Network analysis

A first insight on the network structure can be given analysing the degree distributions for the three models. For an undirected graph, the *degree centrality* of a node is simply the number of neighbors linked

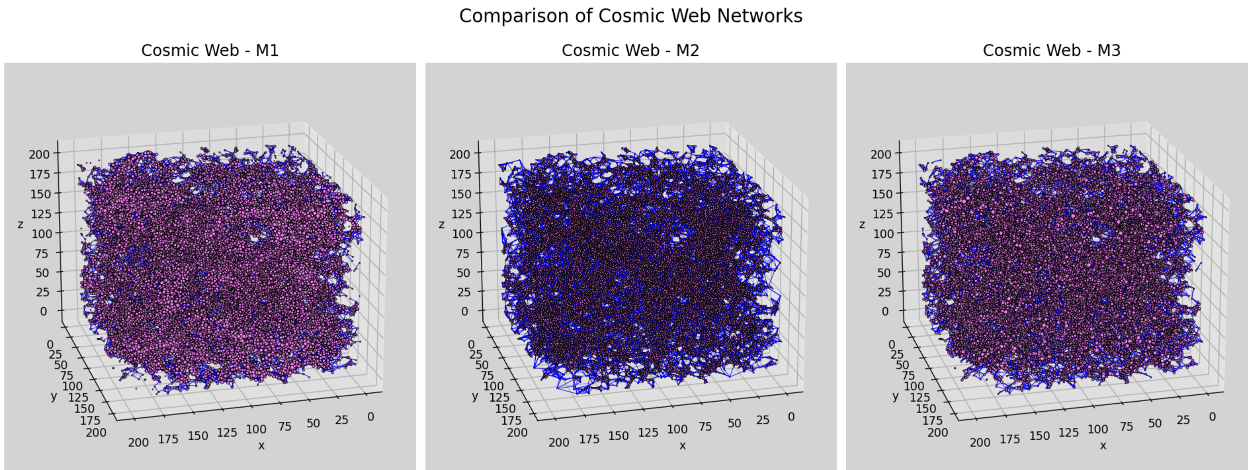

Figure 2: Networks built with the different models.The size of each node is proportional to its degree for the undirected graph M1 and to its in-degree for the directed graphs M2, M3

to it. For directed graphs, we can define the *in-degree* and *out-degree* measures: they count, respectively, the number of edges that point toward and away from the given node. Since the out-degree for models M2 and M3 is fixed, we consider only the in-degree. With this distinction in mind, from now on we will use the denomination DC to refer indiscriminately to the degree for the model M1 and the in-degree for models M2 and M3.

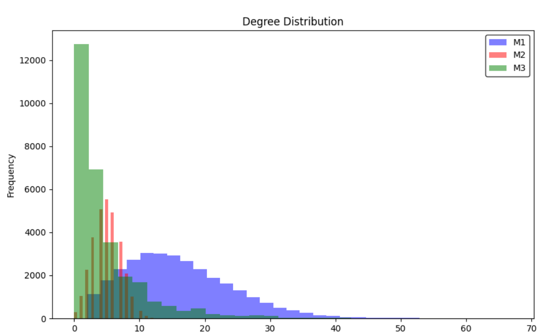

Figure 3: Plot of the degree distribution for the three models.

As we can see from Fig. 3 the distributions are not Poissonian but rather *right-skewed* with a tail of high-degree hubs, indicating that the data we are looking at are not the outcome of a random distribution but of a real, self-organized network [10].

4.1 Parameters justification

We now justify the choice we made for the defining parameters of the three network models. In doing so, we partially follow the approach of B. C. Coutinho et al. (2016)[5]. We study how the *average clustering coefficient* $\langle C \rangle$ and the size of the *largest strongly connected component* S_g (or simply the giant component for the undirected M1 model) change as a function of the parameter.

For undirected and unweighted graphs, the clustering coefficient for a node j is defined as the fraction of possible triangles through that node that exist:

$$c_j = \frac{2y_j}{k_j(k_j - 1)}, \tag{1}$$

with y_j the number of triangles through the node j and k_j its degree. For directed graphs the clustering coefficient is similarly defined in terms of directed triangles:

$$c_j = \frac{y_j}{2(k_j^{tot}(k_j^{tot} - 1) - 2k_j^{\leftrightarrow})}, \tag{2}$$

with k_j^{tot} sum of in- and out-degree, $k_j^{\leftrightarrow}$ the number of bilateral edges between i and its neighbors (i.e., the number of nodes j for which both an edge $i \rightarrow j$ and an edge $j \rightarrow i$ exist) [11]. The average clustering coefficient $\langle C \rangle$ is then simply the average of all cluster coefficients c_j.
For directed graphs, we say that the nodes i and j are strongly connected if there exists a path from i to j and viceversa. Then, a strongly connected component is a maximal subset of vertices such that there is a directed path in both directions between every pair in the subset [10]. This definition does not exist for undirected networks, for which we simply talk about the giant connected component.

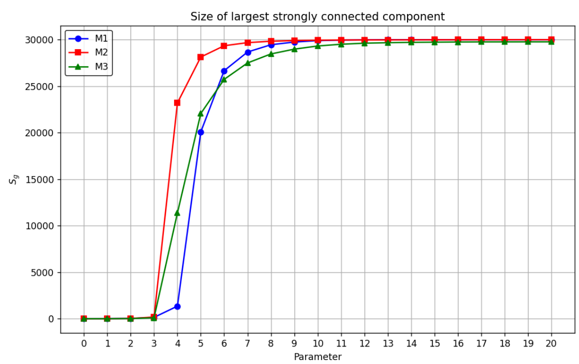

Figure 4: Plot of the size of the largest strongly connected component for the networks as a function of the defining parameter.

From Fig. 4 we see that the largest strongly connected component clearly emerges at $l = 8$ for M1 and at $k = 5$ for M2 and M3 and from Fig. 5 we also see that these values lead to a high average clustering coefficient. Therefore, we use these values for the parameters defining the structure of the different networks. Moreover, we note that the value we found for k is close to the value $k = 4$ found in [5], where it was also observed that the giant strongly connected component emerges at $k = 4$ for the model analogous to our model M2 even for different values of z, meaning that this model can be applied at different redshifts and number of nodes without the need to adjust its parameters, thus reflecting a more fundamental property of how galaxy are distributed in the Universe.

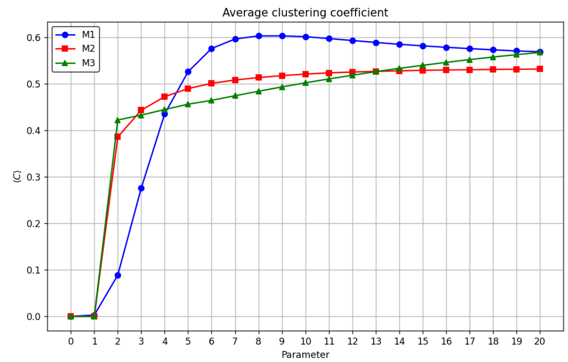

Figure 5: Plot of the average clustering coefficient for the networks as a function of the defining parameter.

4.2 Small-worldness

A further analysis can be made on the concept of a *small-world network*. A network is said to be small-world in nature when the average clustering coefficient $\langle C \rangle$ of the network exceeds the randomly expected one $\langle C_r \rangle$, $\langle C \rangle > \langle C_r \rangle$, and the average path length $\langle l \rangle$ is close or smaller than the randomly expected one $\langle l_r \rangle$, $\langle l \rangle \lesssim \langle l_r \rangle$ [10]. The results for the three models are presented in Tab. 1. Although all networks satisfy the first condition on the clustering coefficient, none of them satisfy the second condition: the average shortest path length of each network largely exceeds the random value, leading to the fact that the Cosmic Web is necessarily a large world, a result in agreement with other works [12].

Network	$\langle C \rangle$	$\langle C_r \rangle$	$\langle l \rangle$	$\langle l_r \rangle$
M1	0.6026	0.0006	26.4734	3.9739
M2	0.4894	0.0002	49.0349	6.5985
M3	0.4559	0.0001	43.5594	6.6415

Table 1: Network metrics for the tree models: $\langle C \rangle$ average clustering coefficient of our network, $\langle C_r \rangle$ average clustering coefficient of random one, $\langle l \rangle$ shortest average path length of our network, $\langle l_r \rangle$ shortest average path length of random one.

4.3 Components of the Cosmic Web

We have said that the structure of the Cosmic Web can be divided into four main components: clusters, filament, walls and voids. The identification and isolation of features and objects in the cosmic matter distribution is essential for understanding the nature of structures which form in the Universe and provides an important link between observation and theoretical models.

Galaxy clusters are the most massive objects in the universe, gathering approximately 4% of the mass of the whole universe. They concentrate near the interstices of the Cosmic Web, with nodes forming a recognizable tracer of the cosmic matter distribution. The richest clusters contain many thousands of galaxies within a relatively small volume of only a few Megaparsec size, but they are first and foremost dense concentrations of dark matter.

Between clusters, galaxies organize themselves in huge *filamentary* or *wall-like* structures. The former define the connective structure, linking higher density regions. The latter describe how galaxies assemble around voids, regions that are devoid of any galaxy.

Voids have a size that can go from 20 to $50h^{-1}Mpc$, they are in general roundish in shape and occupy the major share of space in the Universe.

Using the tools of network analysis, we can trace out clusters, walls and voids from our data evaluating the *degree centrality*, or simply degree. To test its validity, we compare the results obtained with this classification method to the one obtained using the *Voronoi tessellation* [13].

Both Voronoi tessellation and DC are local density measures. The difference is that the Voronoi density is determined by the geometric configuration of neighboring galaxies while the DC method depends on how the network is constructed.

4.3.1 Voronoi tessellation

The Voronoi tessellation corresponds to the creation of a set of mutually disjoint polyhedra, each of which is described by the part of space closest to the point considered compared to all the others points. It is immediate to notice how there is a clear analogy between the constituent elements of the Voronoi tessellation and the various structures that make up the Cosmic Web.

To classify the nodes in one of the different constituent components of the Cosmic Web we use, as a measure, the volume of the cells: a large volume means that the node is in a scarcely populated region whereas a small volume means that the node is surrounded by many neighbors, therefore we assign large volumes to voids and small volumes to clusters, while cells with in-between volumes are assigned to walls. Since the volume thresholds are not defined *a priori* we choose them so that we reproduce the volume distribution found in most works on the Cosmic Web ([1],[14]). These are:

- voids: 77%;
- walls: 18%;
- clusters: $> 0.1\%$.

We have to keep in mind that in our work we don't classify filaments using the DC measure, therefore the leftover volume fraction of the filaments is absorbed by the other components.

The Voronoi tessellation classification method produces the distribution in Fig. 6.

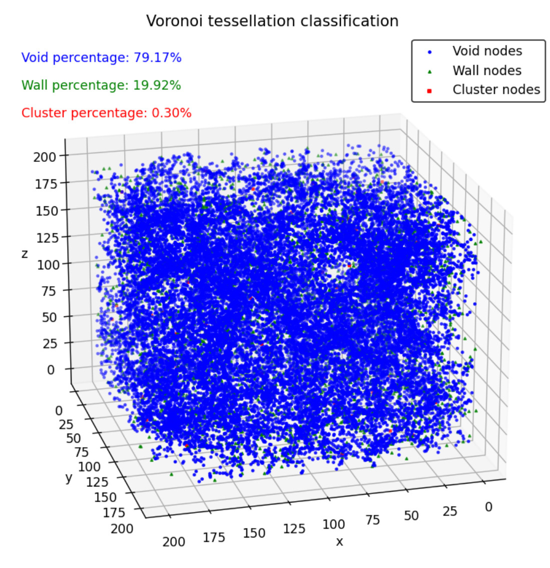

Figure 6: Distribution of nodes in clusters, walls and voids with Voronoi tessellation method.

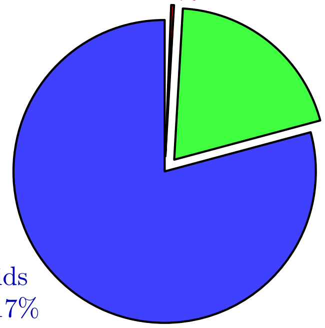

Figure 7: Pie chart of the distribution of nodes in clusters, walls and voids with Voronoi tessellation method.

Because in literature the Voronoi tessellation is largely used we have taken it as reference to optimize the parameters used in the DC method for the various models described in the previous section.

4.3.2 Degree centrality

Given the definition of degree or degree centrality previously stated, we proceed now on using it for a deep investigation of the network topology.
A low DC means that the node is scarcely connected whereas a high value means that the node has many connections. Therefore, as previously done in [15], we classify the nodes as voids, clusters and walls according to their DC value as follows:

- void: DC $< y$;
- wall: $y \leq$ DC $\leq x$;
- cluster: DC $> x$.

The boundaries between the three topological structures are left free in order to investigate which values maximize the agreement between the DC classification method and the Voronoi tessellation method. In Fig. 8 we see how the percentage of matching components between the two classification methods changes, for each model, as a function of the free parameters y and x. We note that, at fixed y, the function is almost insensible to variations in x: this is because the parameter x marks the boundary between walls and clusters and, due to the small fraction of clusters, only nodes with a large DC value are classified as such. What is more important is the dependence on y, which marks the boundary between voids and walls, the components with the largest fraction. In Fig. 9 we observe a side view of the plot: we clearly see that all three distributions reach a plateau after a given y value. In Tab. 2 we report the maximum percentages of matching components between DC and Voronoi tessellation methods for the different networks and the values of the free parameters y and x for which the percentages are maximized.

Network	Maximum percentage	y	x
M1	79.56	49	61
M2	79.65	14	50
M3	79.49	49	60

Table 2: Maximum percentage of matching components between DC and Voronoi tessellation methods for the different networks and values of the free parameters y and x for which the percentage is maximized.

We can conclude that the DC classification method here implemented is, to a good approximation, a valid method for all the three network models used, reproducing to a good degree the distribution of nodes in voids, walls ad clusters as the Voronoi tessellation method.

5 Conclusions and outlook

In this article we used the Complex Network framework to study the Cosmic Web. We constructed the network using three different models: M1, undirected and without constraints on the number of edges, M2 and M3 directed and with a fixed out-degree. Firstly we have outlined the degree distribution that has resulted to be right-skewed as expected from a real network similarly to the internet or the World Wide Web. Then, in order to identify the values of the defining parameters which managed to reproduce the more fundamental aspects of the Cosmic Web, we used the size of the largest strongly connected component and the

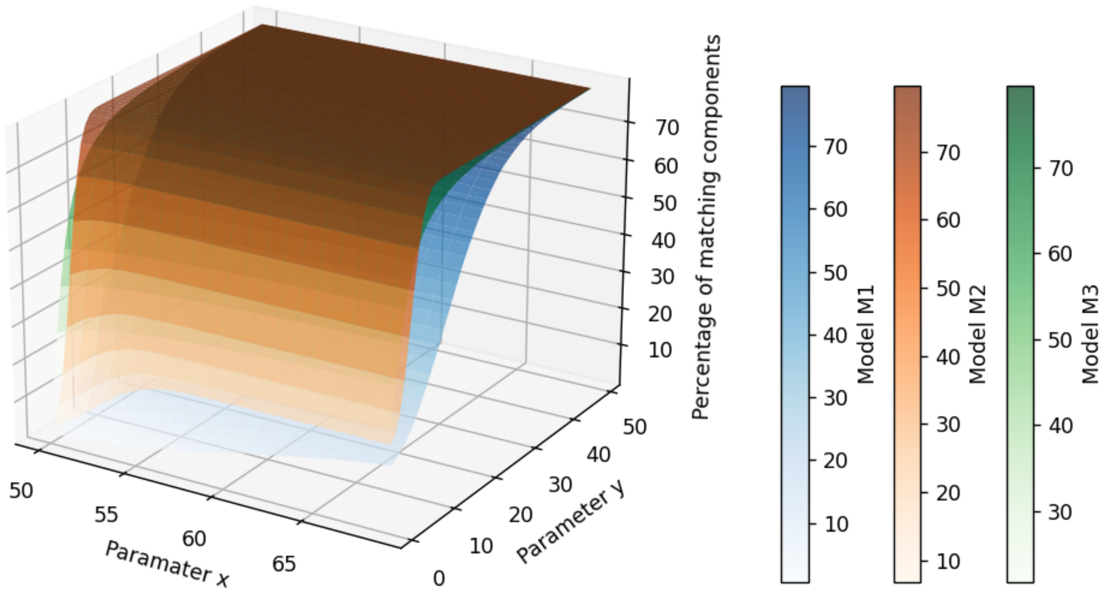

Figure 8: Percentage of matching components between DC and Voronoi tessellation methods.

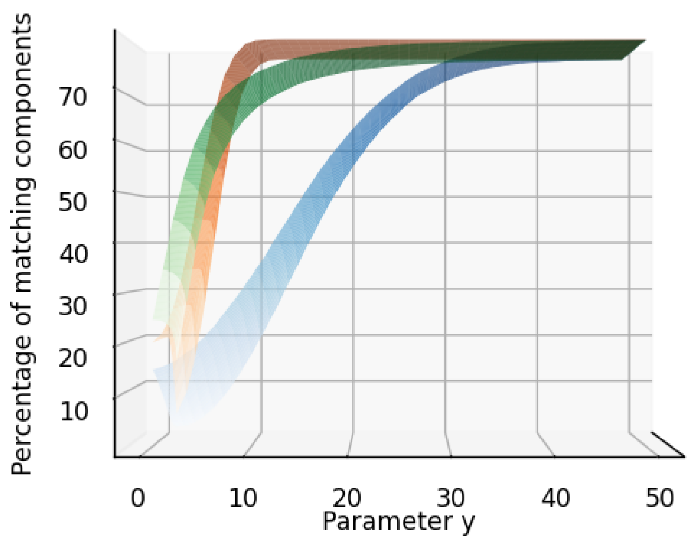

Figure 9: Side view of percentage of matching components highlighting the dependence on the y parameter.

average clustering coefficient of the network for each model.
The results obtained, even if they come from other data, resemble the ones obtained by Coutinho et al. in [5], especially for what regards the M2 model in the value at which the giant component emerges and so its universality for the application to other redshifts. Moreover, we verified that our models correctly produce networks which are not small-world in nature, a well established result for networks based on the Cosmic Web.
Finally, we implemented a method to classify the nodes in voids, walls and clusters based on the degree centrality for the undirected network M1 and on the in-degree centrality for the directed networks M2 and M3. In order to test the validity of the DC classification method, we compared the nodes distribution thus obtained with the well known Voronoi tessellation method, finding a good agreement between the two distributions. This means that our classification method manages to reproduce, to a good approximation, the expected fraction of voids, walls and clusters. As a further development, it could be interesting to study different classification methods from the one here implemented, based on other centrality measurements, e.g. the betweenness centrality or the closeness centrality. Moreover, to better study the validity of the chosen classification method, it could be compared with different algorithms other than the Voronoi tessellation, such as the NEXUS+ algorithm [14] or the MST one based on percolation [16].

Bibliography

[1] Rien van de Weygaert and J. R. Bond. "Observations and Morphology of the Cosmic Web". In: *Lect. Notes Phys.* 740 (2008), pp. 409–467.

[2] J. H. Oort. "Superclusters". In: *Annual review of astronomy and astrophysics* 21 (1983), pp. 373–428.

[3] Einasto J. and Jôeveer M. "Superclusters and galaxy formation". In: *Nature* 283 (Jan. 1980). DOI: 10.1038/283047a0. URL: https://doi.org/10.1038/283047a0.

[4] M. Davis et al. "A survey of galaxy redshifts. II. The large scale space distribution." In: *Astrophysical Journal* 253 (Feb. 1982), pp. 423–445. DOI: 10.1086/159646.

[5] B. C. Coutinho et al. *The Network Behind the Cosmic Web.* 2016.

[6] Noam I. Libeskind et al. "Tracing the cosmic web". In: *Monthly Notices of the Royal Astronomical Society* 473.1 (Aug. 2017), pp. 1195–1217. DOI: 10.1093/mnras/stx1976. URL: https://doi.org/10.1093%2Fmnras%2Fstx1976.

[7] D. Kelesirs, D. Fotakis S. Basilakos V. Papadopoulou Lesta, and A. Efstathiou. "Detecting and analysing the topology of the cosmic web with spatial clustering algorithms 1: methods". In: *Monthly Notices of the Royal Astronomical Society* 516 (Sept. 2022), pp. 5110–5124. DOI: 10.1093/mnras/stac2444. URL: https://doi.org/10.1093/mnras/stac2444.

[8] Volker Springel. "The cosmological simulation code gadget-2". In: *Monthly Notices of the Royal Astronomical Society* 364.4 (Dec. 2005), pp. 1105–1134. DOI: 10.1111/j.1365-2966.2005.09655.x. URL: https://doi.org/10.1111%2Fj.1365-2966.2005.09655.x.

[9] Planck Collaboration et al. "Planck 2013 results. XVI. Cosmological parameters". In: 571, A16 (Nov. 2014), A16. DOI: 10.1051/0004-6361/201321591. arXiv: 1303.5076 [astro-ph.CO].

[10] M. E. J. Newman. *Networks an introduction.* Oxford University Press, 2010.

[11] Giorgio Fagiolo. "Clustering in complex directed networks". In: *Phys. Rev. E* 76 (2 Aug. 2007), p. 026107. DOI: 10.1103/PhysRevE.76.026107. URL: https://link.aps.org/doi/10.1103/PhysRevE.76.026107.

[12] R. de Regt et al. "Network analysis of the COSMOS galaxy field". In: *Monthly Notices of the Royal Astronomical Society* 477 (Mar. 2018), pp. 4738–4748. ISSN: 0035-8711. DOI: 10.1093/mnras/sty801. URL: https://doi.org/10.1093/mnras/sty801.

[13] Franz Aurenhammer, Rolf Klein, and Der-Tsai Lee. *Voronoi Diagrams and Delaunay Triangulations*. 1st. USA: World Scientific Publishing Co., Inc., 2013. ISBN: 9789814447638.

[14] Marius Cautun et al. "Evolution of the cosmic web". In: *Monthly Notices of the Royal Astronomical Society* 441.4 (May 2014), pp. 2923–2973. ISSN: 0035-8711. DOI: `10.1093/mnras/stu768`. URL: `https://doi.org/10.1093/mnras/stu768`.

[15] Sungryong Hong and Arjun Dey. "Network analysis of cosmic structures: network centrality and topological environment". In: *Monthly Notices of the Royal Astronomical Society* 450.2 (Apr. 2015), pp. 1999–2015. ISSN: 0035-8711. DOI: `10.1093/mnras/stv722`. URL: `https://doi.org/10.1093/mnras/stv722`.

[16] J. D. Barrow, S. P. Bhavsar, and D. H. Sonoda. "Minimal spanning trees, filaments and galaxy clustering". In: 216 (Sept. 1985), pp. 17–35. DOI: `10.1093/mnras/216.1.17`.