

¹ DRAFT VERSION AUGUST 16, 2019

Corresponding author: Felipe Leonardo Gómez-Cortés
fl.gomez10@uniandes.edu.co

² Typeset using L^AT_EX **preprint** style in AASTeX62

³ A Large Scale Structure Void Identifier for Galaxy Surveys Based on the β -Skeleton Graph Method

⁴ FELIPE LEONARDO GÓMEZ-CORTÉS,¹ JAIME E. FORERO-ROMERO,¹ AND XIAO-DONG LI²

⁵ *Physics Department, Universidad de Los Andes*

⁶ *Shcool of Physics, Korean Institute for Advanced Study*

7 ABSTRACT

8 1. INTRODUCTION

9 2. THE ALGORITHM

10 This void identifier is based on the β -Skeleton graph. The input is a set of 3-dimensional points
 11 (galaxy/halo catalog) in the coordinate space with proper distances. This is the Observed Catalog
 12 (**OC**). Once is defined the volume and shape of the **OC**, this space is populated with random points
 13 following a uniform probability density function, this is the Random Catalog (**RC**). The Full Catalog
 14 (**FC**) is made by assembling the OC and RC. The β -Skeleton graph is built over the FC without any
 15 discrimination connecting observed and random points, the graph is stored as **fcBSkel**. From the
 16 **fcBSkel**, all the observed points and their connections are deprecated; it means that some neighbour
 17 random points are rejected. The remaining random points are connected only with other random
 18 points, they conform the voids.

19 This method has two parameters: β and the ratio between random points and observed points.

20 2.1. Testing methods with the Void Toy Models

21 The Li's β -Skeleton library ([LSSCode 2014](#)) is written in Fortran 90. The calculation of the 1-
 22 Skeleton graph with $\sim 10^4$ points in a regular laptop Linux machine (Core i5, 2nd gen.) takes 192
 23 seconds. Input files of three dimensionless columns (x, y, z): \mathbb{R}^3 points.

24 Observational Toy Model Catalogs where created to test if the algorithm was viable.

25 The first toy model catalog was generated with a spherical void into inside a cubic box (100 arbitrary
 26 units of lenght) filled with N_{obs} random uniformly distributed points.

27 2.1.1. The weak method

28 A first attempt was an algorithm based on the comparison of two β -Skeleton graphs ($\beta_1 < 1$),
 29 ($\beta_2 \geq 1$), (fig. 1). This method was able to detect multiple spherical voids, but was unable to
 30 detect the whole enclosing surface of elliptical voids, the second testing toy model.

31 2.1.2. The robust method

32 The second and definitive algorithm relies on random points filling voids. The algorithm sucessfully
 33 recognized the spherical void, and also identified some other underdense regions, figure 2-left. A limit
 34 resolution arises, related to the number of random points, it was shown finding small spherical voids
 35 placed in different locations, figure 2-right.

36 The robust method cand find irregular voids. The last void toy model used overlapping ellipsoidal
 37 voids with semiaxes ratios 1:0.7:0.5, different orientations and major semiaxes of 40, 30, 30 and 20
 38 arbitrary units, generating a big irregular void. The OC and a slice is shown at figure 3.

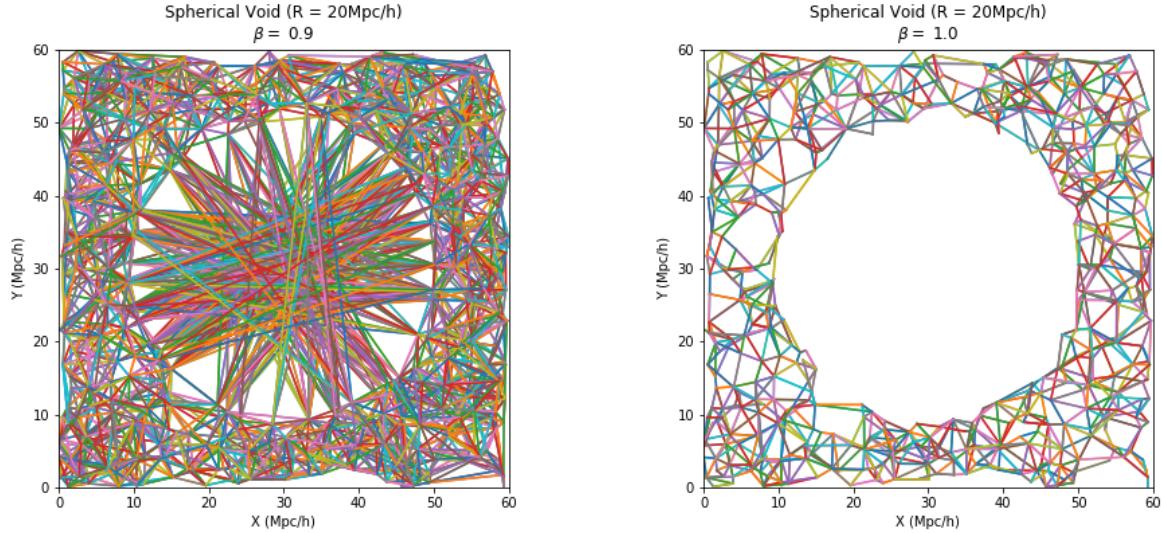


Figure 1. Weak method: Long connections dissapear in the β -Skeleton graph as β increases. Left: Connections with $\beta = 0.9$. Right: Connections with $\beta = 1.0$. The low value of β gives the graph long connections across the spherical void. They dissapear with $\beta = 1.0$, known as the Gabriel Graph.

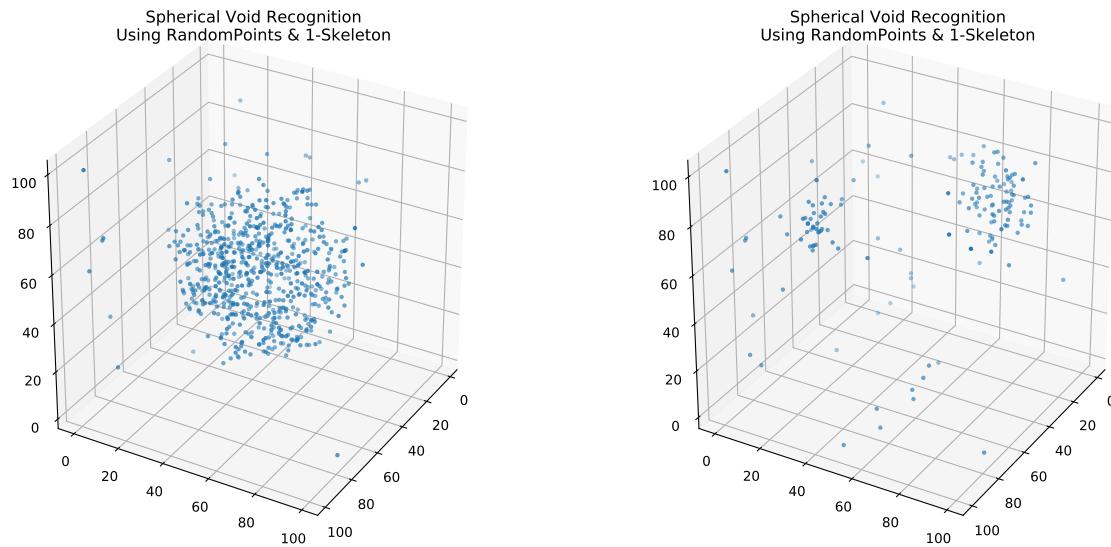


Figure 2. The robust method identifies voids by populating the volume of the observed catalog with random points. Parameters: $\beta = 1$ and $n_{rnd} = 1$. Left: Random Points identified as “Void Points”. They are points from the Random Catalog connected by the 1-Skeleton only to other Random Catalog points. $N=5000$ points, BoxLength = 100 units and void radius $R=37$ units. Right: Points identified as “Void Points”. This tests aims to check the resolution of the algorithm. Some scattered micro-voids appear. The two big voids are recognized ($R=15,20$), but small structures doesn't appear ($R=5,10$). $N=5000$, BoxLength = 100, voids radius $R = [5, 10, 15, 20]$.

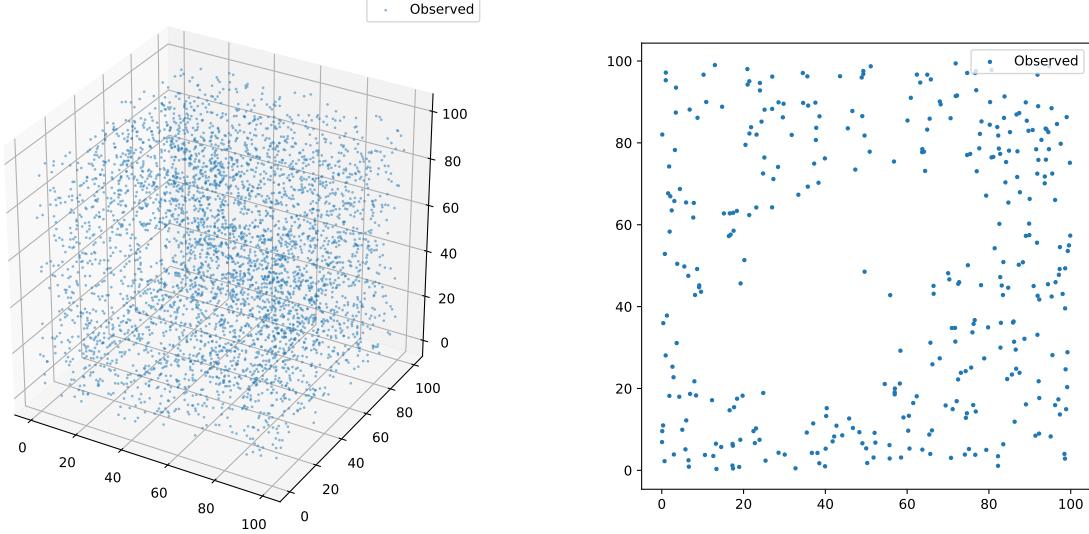


Figure 3. A Irregular Void in the testing toy model. Observed points (fake galaxies) in blue. Left: 3D scatter-plot of the full sample of OC. Right: A slice of the OC at $z = 50 \pm 5$ arbitrary units.

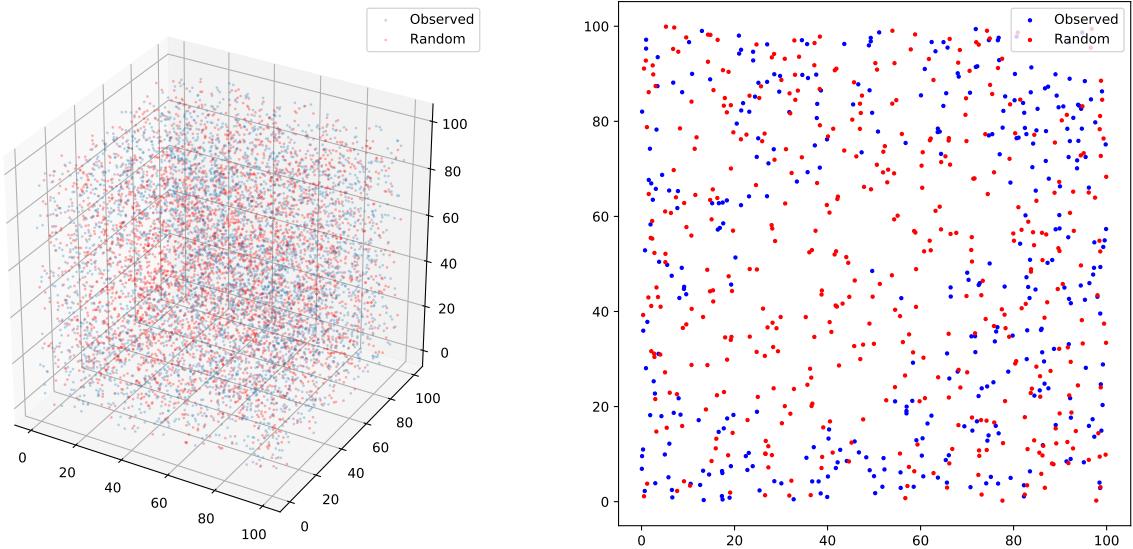


Figure 4. The Random Catalog (red) of points populates the whole volume of the (blue) OC. In this case, we have the same number of Random Points and Observed Points. Left: 3D scatter-plot of the OC plus RC. Right: A slice of the OC plus RC at $z = 50 \pm 5$ arbitrary units.

39 The Random Catalog was created using the same number of random points as the observed
 40 points. $N_{rnd} = N_{obs}$ (Fig.4). It follows to the recognition of the main structures (ellipsoidal voids
 41 overlapping as a bigger one) plus small unexpected structures. Those are clue of the relevance of
 42 the ratio $n_{rnd} = N_{rnd}/N_{obs}$ as free parameter of the algorithm. Having $n_{rnd} > 1$ means a higher
 43 resolution in the shape of the identified voids, but also smaller voids are detected in the catalog; are
 44 they artifacts or true voids? The answer deppends on the void definition itself. The initial test is
 45 performed with $n_{rnd} = 1$.

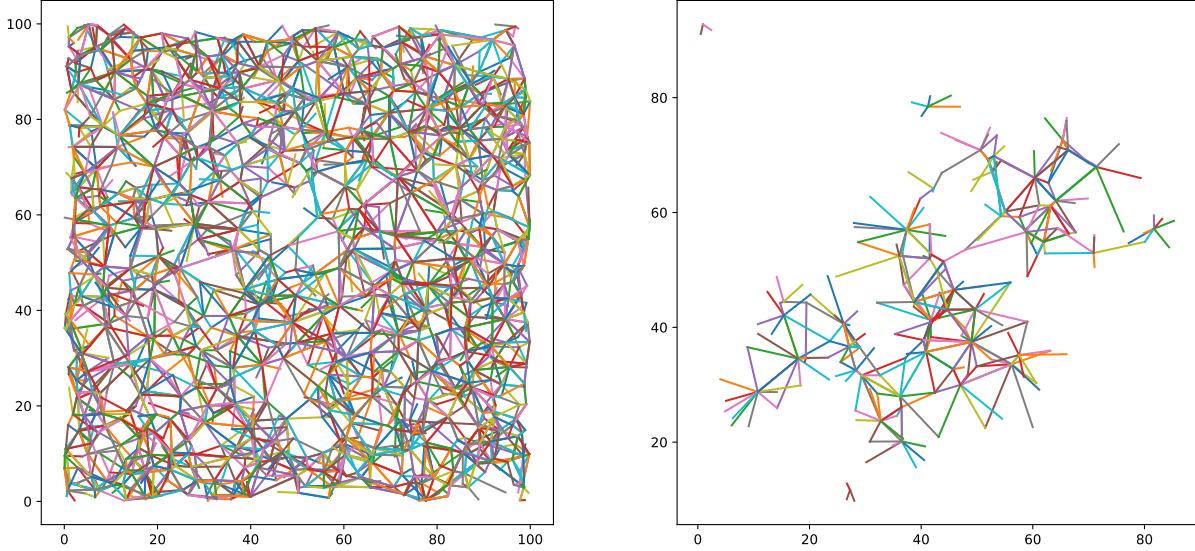


Figure 5. A slice ($z = 50 \pm 5$ arbitrary units) of the β -Skeleton of the full catalog (left). After removing the observed points and their connections, the remaining points are the void points (right).

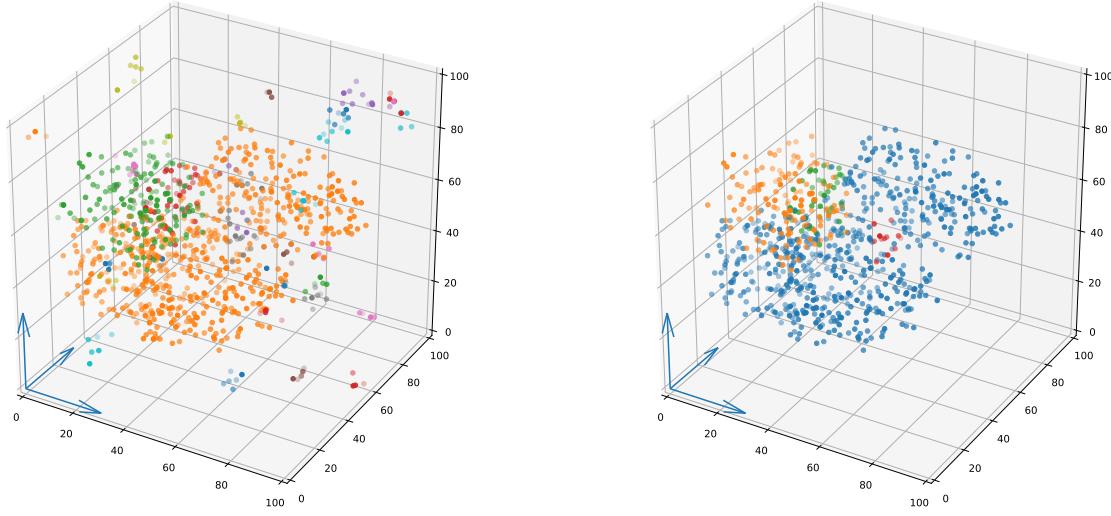


Figure 6. Irregular voids Voids found in the OC. This algorithm can indentify irregular structures; a giant leap ahead the previous algorithm. The irregular voids are made of overlapping ellipsoidal voids, with semiaxes proportion 1:0.7:0.5 as the literature suggests. N=5000, BoxLength = 100, the voids major semiaxes where chosen as R= 40, 30, 30 and 40 arbitrary units. Left: A total of 41 voids are identified, all of them with at least one random point. To take into account, only four toy ellipsoidal voids where placed in the OC, two of them overlapping. Right: Voids with 10 or more random points.

2.2. Tagging Voids

46 So far, the method can isolate the true void points: RC points that are not connected with the
 47 OC according to the β -Skeleton graph. The identification of structures uses the same graph, using a
 48 recursive search of connected neighbours; a recursive search of friends of friends.
 49

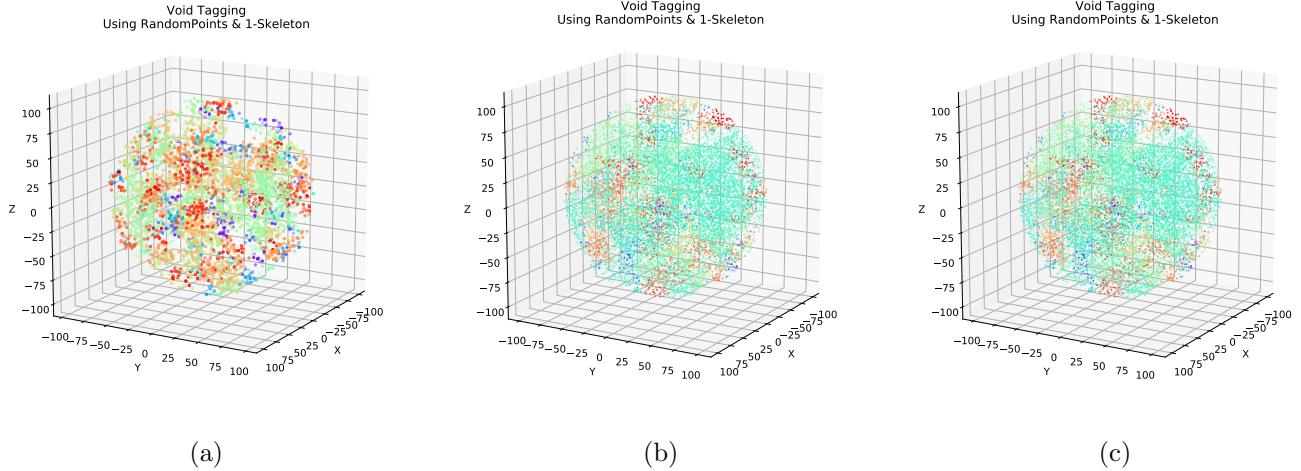


Figure 7. Tagging Abacus-Cosmos Voids. This is the first test over “real” data, not a toy model. The LSS shows highly dense packed halos, where voids have almost none of them. The idea is to increase the ratio of Random Points over Catalog points to study (in later weeks) how void detection can be affected. (a) Ratio 1:1, (b) ratio 2:1 and (c) ratio 3:1. The catalog is a spherical set of points, with radius $100\text{Mpc}/h$, N halos = 9981, N random points = 10000 (a), $2N$ (b) and $3N$ (c).

3. FINDING VOIDS IN SIMULATIONS

3.1. *Abacus-Cosmos Simulation*

52 From a snapshot of the Abacus-Cosmos simulation with fiducial cosmology, a set of halos is selected,
 53 within a sphere of radius $R = 100\text{Mpc}/h$ and a M_{cut} given to get $\sim 10^4$ halos.

The Random Catalog is generated, taking care of having a uniform space distribution. A common error is to generate the radius using a uniform distribution from 0 to 1 and scaling by a factor R , but must be taken instead the cubic root of the random number between 0 and 1, then multiplied by R . Another error is generating the angles again, taking a random number between zero and one, then scaling by 2π . This error can be avoided by generating a random number between 0 and 1 as the cosine of θ , then calculating the angle as the arc-cosine. The unitary vector can be calculated using a gaussian distribution for x, y and z (centered at zero with the same deviation for them three). The norm is calculated as usually, by taking the square root of the sum of the coordinates, then dividing each coordinate by the norm of the vector.

⁶³ The algorithm ran over the “real” dataset using the ratios 1:1, 2:1 and 3:1 for the number of
⁶⁴ random points vs. catalog points.

3.2. Running over a single spherical cut

66 A kind of percolation phenomenon is appreciated at figure 7. Having a similar number of Random
67 Points and Observational Points (a) conducts to sucessfull void identification, while a higher ration
68 2:1 (b) and 3:1 (c)

3.3. Running over 64 spherical cuts

3.4. Running over Corner Plots

4. FINDING VOIDS IN GALAXY REDSHIFT SURVEYS

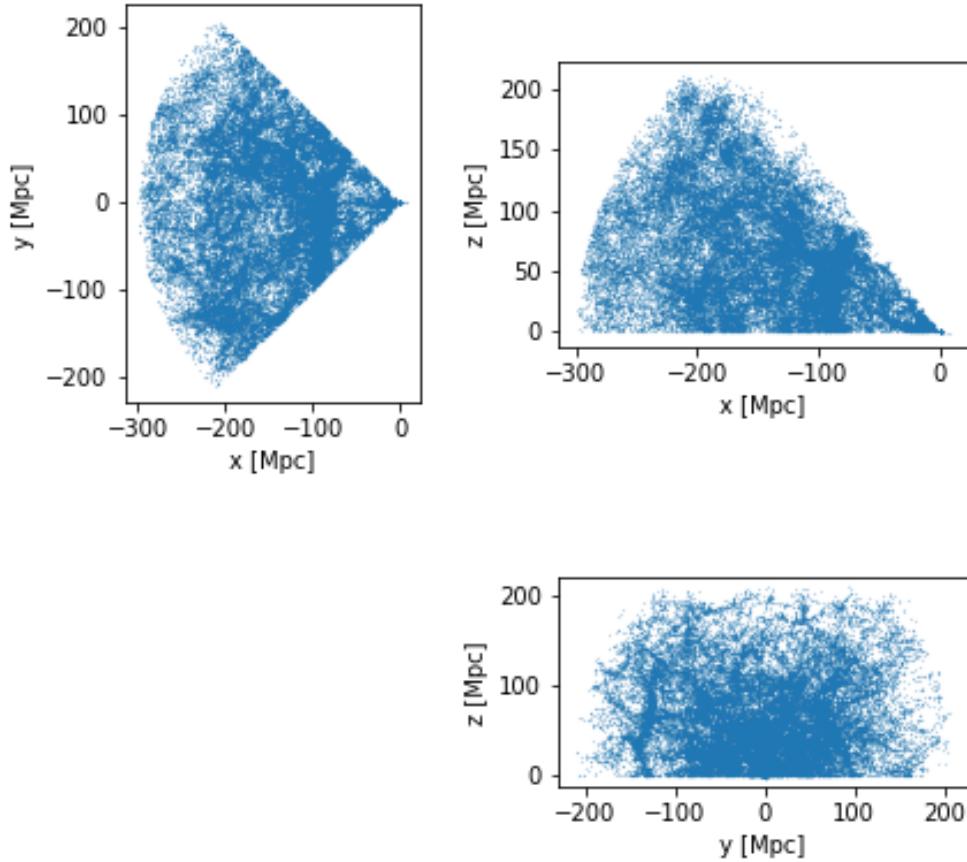


Figure 8. Dataset from SDSS. Comoving distances are in Mpc, calculated from the redshift measurements using the module “cosmology” from the library “astrypy” with Planck15 cosmological parameters.

4.1. SDSS Planck 2015 data

The SDSS has ~ 57000 galaxies in the truncated cone section from RA 0 to 50 deg, and DEC from -40 deg to $+40$ deg. Using Planck-2015 cosmology, reshift was converted into the range of distance from 0 to 300 Mpc. (figure 8).

The void size density function is calculated as the probability to find a void of a given radius in to the total volume. There is a discrepancy between our density function and Platen (2008) results, maybe they did not normalize by the volume. Figure 9.

The void ellipticity is calculated as $\epsilon = 1 - c/a$, with a the biggest semi-axe, and c the smallest semi-axe. They two graphs seem to have similar values. The probability density is not normalized by the volume. Figure 10.

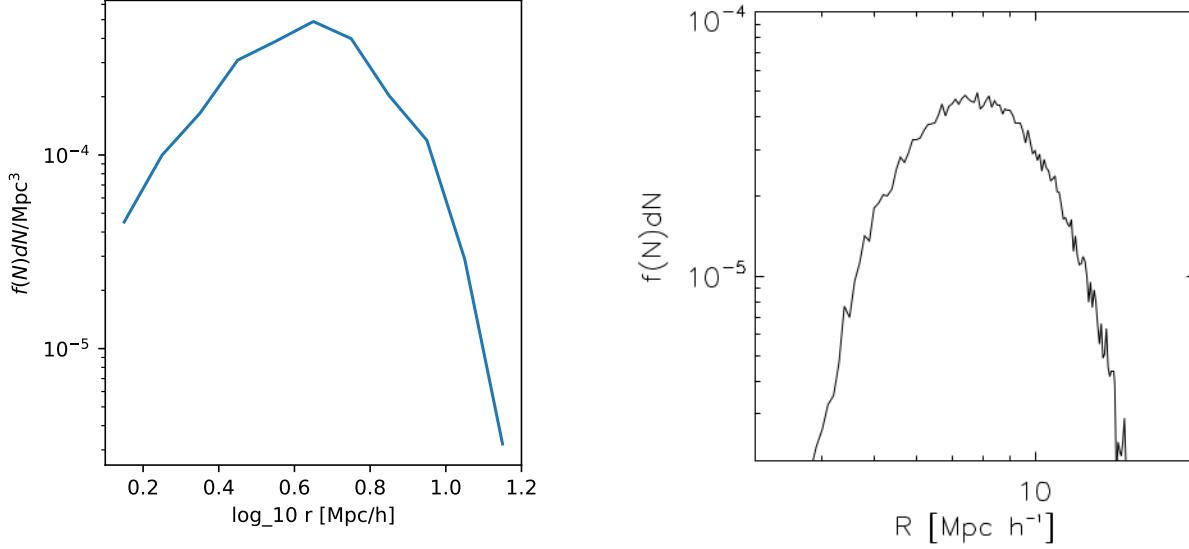


Figure 9. Void size density function. Left, our analysis. Right, Platen et al (2008).

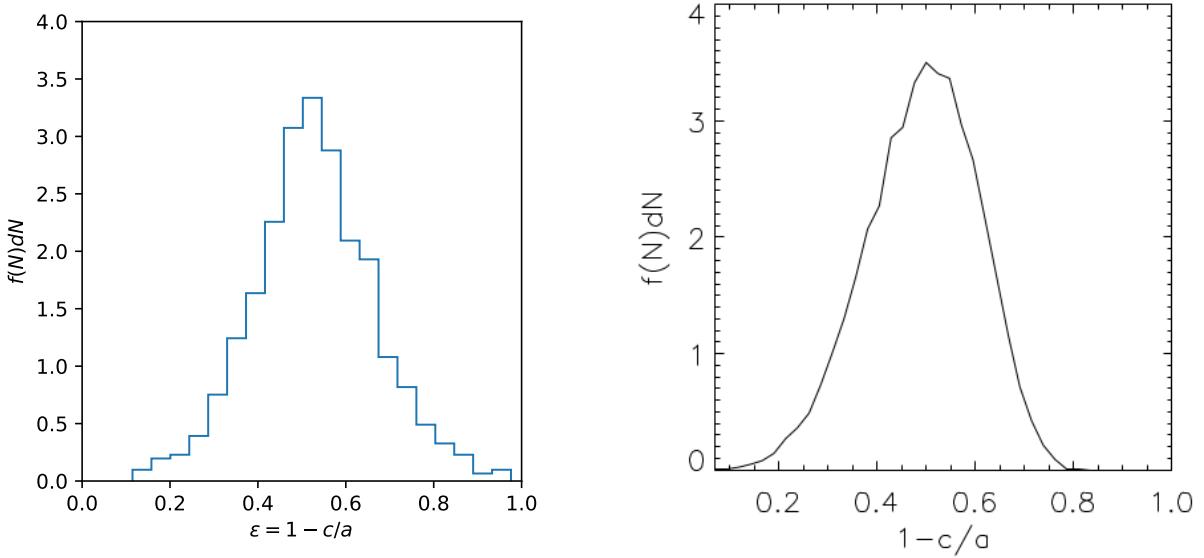


Figure 10. Void Ellipticity. Left, analysis of voids found with our algorithm. Rigth, Platen et al (2008).

83 The last comparison was made using the two axis ratios to find how prolate, oblate or close to
84 spheroid. Our results are pretty close to Platen results (2008).

85 *Software:* LSSCode ([LSSCode 2014](#))

REFERENCES

- | | |
|--|--|
| 86 Li, Xiao-Dong 2014, LSSCode
87 https://github.com/xiaodongli1986/LSS_Code | 88 Astropy Collaboration, Robitaille, T. P., Tollerud,
89 E. J., et al. 2013, A&A, 558, A33 |
|--|--|

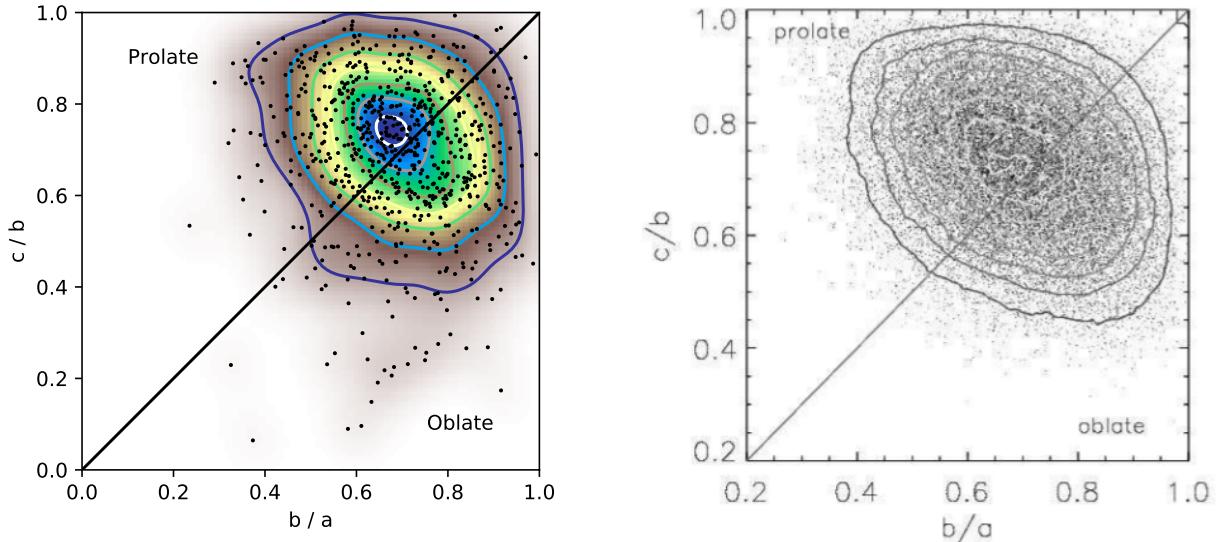


Figure 11. Scatter diagram of two axis ratios. In the approximation of the voids as ellipsoidal forms, the semi-axes are labeled, from shorter to larger, as a , b and c . A perfect spherical void should be placed at the right top corner of the diagram.