

Improving SDSS cosmological constraints through β -skeleton weighted correlation functions

Fenfen Yin^{1,2}, Jiacheng Ding,¹ Limin Lai,^{1,3} Wei Zhang,¹ Liang Xiao¹, Zihan Wang,¹
Jaime Forero-Romero^{4,*}, Le Zhang,^{1,5,6,†} and Xiao-Dong Li^{1,5,6,‡}

¹*School of Physics and Astronomy, Sun Yat-Sen University, Zhuhai 519082, China*

²*School of Data Science, Tongren University, Tongren 554300, China*

³*Department of Astronomy, Shanghai Jiao Tong University, Shanghai 200240, China*

⁴*Departamento de Física, Universidad de los Andes, Cra. 1 No. 18A-10 Edificio Ip, CP 111711, Bogotá, Colombia*

⁵*Peng Cheng Laboratory, No. 2, Xingke 1st Street, Shenzhen 518000, China*

⁶*CSST Science Center for the Guangdong–Hong Kong–Macau Greater Bay Area, SYSU, Zhuhai 519082, China*



(Received 25 March 2024; accepted 28 May 2024; published 24 June 2024)

The β -skeleton approach can be conveniently utilized to construct the cosmic web based on the spatial geometry distribution of galaxies, particularly in sparse samples. This method plays a key role in establishing the three-dimensional structure of the Universe and serves as a tool for quantitatively characterizing the nature of the cosmic web. This study is the first application of β -skeleton information as weights in mark weighted correlation functions (MCFs), presenting a novel statistical measure. We have applied the β -skeleton approach to the CMASS NGC galaxy samples from SDSS BOSS DR12 in the redshift interval $0.45 \leq z \leq 0.55$. Additionally, we applied this approach to three COLA cosmological simulations with different settings ($\Omega_m = 0.25$, $\Omega_m = 0.31$, $\Omega_m = 0.4$) for comparison. We measured three MCFs, each weighted by (i) the number of neighboring galaxies around each galaxy, (ii) the average distance of each galaxy from its surrounding neighbors, and (iii) the reciprocal of the average distance of each galaxy from its surrounding neighbors. By comparing measurements and calculating corresponding χ^2 statistics, we observe high sensitivity to the cosmological parameter Ω_m through a joint analysis of the two-point correlation and three MCFs.

DOI: [10.1103/PhysRevD.109.123537](https://doi.org/10.1103/PhysRevD.109.123537)

I. INTRODUCTION

According to the principles of cosmology, the Universe is considered homogeneous and isotropic at a large scale. However, the combined influence of gravity and the ongoing expansion of the Universe results in a complex and nonlinear structure. In terms of spatial configuration, the unique and intricate web structure formed by the anisotropic collapse of gravity is referred to as the “cosmic web” [1–6]. Topologically, based on the dimensions of gravitational collapse, the cosmic web is typically categorized into knots, filaments, sheets, and voids.

The study of the structure of the cosmic web relies on large-scale redshift surveys, and currently, there are a number of large-scale observational projects, such as the Sloan Digital Sky Survey (SDSS) [7–18], the 2-degree Field Galaxy Redshift Survey (2dFGRS) [19], the 6-degree

Field Galaxy Redshift Survey (6dFGRS) [20,21], and the WiggleZ Dark Energy Survey [22].

The ongoing stage-IV surveys, such as the Dark Energy Spectroscopic Instrument (DESI) [23,24], are expected to further deepen our understanding of the Universe. They possess the capability to observe wider and deeper fields of view with higher resolution, allowing for the collection of more accurate and extensive data and providing additional information about the nonlinear scale structure of the Universe.

The China Space Station Telescope (CSST) [25], scheduled for launch in 2025, is a forthcoming stage-IV galaxy survey. It will conduct observations across 17,500 square degrees of the sky, capturing images in the *ugriz* bands with a spatial resolution comparable to that of the Hubble Space Telescope (HST).

The exploration of the cosmic web has consistently been a significant topic in the study of the Universe. Since early maps of the Universe confirmed the existence of such webs through galaxy redshift measurements, there has been a continuous effort over the past 40 years to find a consistent and stable method to define these weblike

*je.forero@uniandes.edu.co

†zhangle7@mail.sysu.edu.cn

‡lixiaod25@mail.sysu.edu.cn

structures. Currently, the commonly used analytical tools for studying cosmic webs such as β -skeleton [26], FoF [27], density-based techniques [28–30], T-web [31,32], and V-web [33,34].

In 2019, the β -skeleton analysis was employed for the first time to characterize the cosmic web, by analyzing both observational data and accompanying simulations. The observational data utilized the BOSS DR12 CMASS [35] galaxy sample, and the BigMultiDark N-body simulation conducted under the Planck 2015 Cosmology [36] served as a test sample. By varying the β value and examining the many different scenarios such as redshift evolution and redshift space distortion, it is observed that the number of skeleton connections decreases with increasing β , resulting in a sparser cosmic web. For instance, when $\beta = 1$, numerous connections are generated, irrespective of whether the nodes are located in voids or filaments. However, with setting $\beta = 3$, the resulting cosmic web closely resembles that of real observations, indicating a nearly perfect match in the number of connections to the actual galaxies. Finally, at $\beta = 10$, only small and relatively isolated compact groups of galaxies are identified and connected.

Moreover, [37] explored the information theory entropy of a graph as a scalar to quantify the cosmic web. The findings suggest that entropy can be used as a discrete analogue of scalars used to quantify the connectivity in continuous density fields. The simplicity of computing graph entropy allows its application to both simulations and observational data from large galaxy redshift surveys. This new statistic offers a complementary approach to other topological or clustering measurements. More recently, using data from the IllustrisTNG simulation [6,38–40] explored a fast machine learning-based approach to infer the underlying dark matter tidal cosmic web environment of a galaxy distribution from its β -skeleton graph. The β -skeleton is more suitable for capturing the relatively sparse observed cosmic web compared to traditional T-web and V-web methods. It achieves this by straightforwardly establishing the three-dimensional structure of the Universe based on the spatial geometry of galaxies. In addition, this method overcomes the need for artificially setting a smoothing scale, unlike other methods. It adapts to the density of the environment, generating mathematical topology that traces the details of galaxy distribution in both high- and low-density regions simultaneously. This approach provides a more convenient and feasible way to construct cosmic webs.

The current structure of the Universe has evolved from early small density fluctuations. Through the combined effects of gravitational evolution and cosmic expansion, it gradually formed into the complex structure of cosmic webs observed today, displaying various morphologies at different scales. The traditional and commonly used methods for analyzing the statistical properties of the large-scale

structure (LSS) of the Universe include the two-point correlation function (2PCF) or the power spectrum in Fourier space. However, these methods essentially measure the Gaussian content of the density field. With the non-linear evolution caused by gravitational collapse, the significance of non-Gaussianity increases over time. While higher-order statistics, such as the three-point correlation function [41,42] and the four-point correlation function [43], can capture non-Gaussian information, they face challenges in terms of visual interpretation and efficient calculation.

Recently, the mark weighted correlation function (MCF) [44], which assigns density weights to different features of galaxies to extract non-Gaussian information on LSS, has proven effective in capturing more detailed clustering information. MCFs are computationally more feasible compared to other statistical methods capable of extracting non-Gaussian information from LSS. By selecting weights based on the α th power of the local density, the weighting amplifies either dense or low-density regions. This method can enhance the distinction between galaxy clusters in dense and sparse regions, resulting in more stringent constraints compared with 2PCF. References [45,46] demonstrate that the joint constraint of the MCFs with different α values significantly improved constraints on cosmological parameters, such as Ω_m and w . Reference [47] presents measurements of MCFs of LOWZ galaxies from SDSS BOSS DR12 to distinguish between the Λ CDM model and $f(R)$ gravity models. Reference [48] employed density gradients as weights and find that the gradient-weighting scheme would produce more robust parameter constraints compared to the density-marked scheme. Furthermore, combining these two schemes yields even stronger constraints than using either one alone.

In this study, we present the first application of the β -skeleton as weights in MCFs, serving as a statistical measure of information about the environment contained in the knots of the β -skeleton. Constructing undirected graph information from the β -skeleton allows for extracting additional information about LSS. We will apply three weighting schemes to MCFs, utilizing undirected graph information derived from the β -skeleton. These schemes involve weighting by (i) the number of neighboring galaxies around each galaxy, denoted as N_{con} ; (ii) the average distance of each galaxy from its surrounding neighbors, denoted as \bar{D}_{nei} ; and (iii) the reciprocal of the average distance of each galaxy from its surrounding neighbors, $1/\bar{D}_{\text{nei}}$.

This paper is organized as follows: Sec. II details the COLA and PATCHY simulations, as well as the Sloan Digital Sky Survey III (SDSS-III) Baryon Oscillation Spectroscopic Survey (BOSS) Data Release 12 (DR12) galaxy sample, which are utilized in the subsequent analysis. In Sec. III, we provide a brief introduction to the definition and basic properties of the β -skeleton,

outlining the methodology for calculating the MCFs on the data. In Sec. IV, we present our findings by applying the MCFs using the weights derived from the β -skeleton quantities. Finally, we summarize our results and draw conclusions in Sec. V.

II. DATA

To investigate the sensitivity of MCF statistics using the properties of the β -skeleton as the weighting scheme for cosmological parameters, we need to compare actual observations with a series of simulations. Our analysis relies on observations of SDSS BOSS DR12 CMASS galaxies. Additionally, the analysis in this work depends on a set of fast simulations generated using the COmoving Lagrangian Acceleration (COLA) method [49]. The galaxy mock catalog accurately reproduces the statistical properties of observed CMASS galaxies by employing a recently proposed scheme [50] that utilizes the subhalo abundance-matching (SHAM) procedure and involves adjusting the COLA simulation parameters. Note that to ensure the accuracy of COLA for our analysis, we have compared various statistics with those derived from GADGET simulations using the same cosmological parameters and found agreement between the two (see the Appendix for details). Additionally, we utilize the Multidark PATCHY simulations (referred to as PATCHY) [51] to estimate the relevant covariance.

A. Observational data

The inclusion of the Baryon Oscillation Spectroscopic Survey (BOSS) is a component of SDSS-III [52–54]. The detection of the characteristic scale imprinted by baryon acoustic oscillations (BAO) in the early Universe is aimed for by the measurement of the spatial distribution of luminous red galaxies (LRGs) and quasars through BOSS [55].

Redshift information for approximately 1.5 million galaxies in a sky region of $\sim 10^4$ square degrees is provided by BOSS, which is divided into two samples: LOWZ and CMASS. The LOWZ sample consists of the brightest and reddest LRGs at $z \leq 0.4$, while the CMASS sample targets galaxies at higher redshifts, many of which are also LRGs. The sky region covered by CMASS NGC galaxies is approximately $\text{R.A.} \in [108^\circ, 264^\circ]$, $\text{Dec} \in [-4^\circ, 69^\circ]$. Our study specifically concentrates on a subset of the BOSS DR12 CMASS NGC galaxies within a redshift range of $z \in [0.45, 0.55]$.

B. COLA simulations and mock galaxy catalogs

COLA,¹ a hybrid approach integrating second-order LPT (2LPT) and N-body algorithms, proves to be an effective solution for simulating dark matter (DM) particles.

Perturbation theory has demonstrated success in describing large scales, allowing the linear growth rate to replace time integration in N-body simulations. COLA leverages this by employing a comoving frame, with observers following trajectories calculated in perturbation theory. Importantly, COLA accurately calculates large-scale dynamics using 2LPT while assigning the resolution of small scales to the N-body code, without demanding on an exact representation of the internal dynamics of halos. Consequently, COLA can trade accuracy at small scales for computational speed without compromising accuracy at large scales.

To assess the performance of our method and its sensitivity to cosmological parameters, we utilize a series of COLA simulations. To examine the dependence on cosmological parameters, three sets of COLA simulations are employed, each adopting different present-day dark matter density: $\Omega_m = \{0.25, 0.31, 0.4\}$, respectively. The simulations are based on a Λ CDM cosmology with the following parameter values fixed: $\Omega_b = 0.048$, $w = -1.0$, $\sigma_8 = 0.82$, $n_s = 0.96$, and $h = 0.67$. These values closely approximate the mean constraint derived from the Planck 2015 results [56]. We ran COLA simulations with a total of 1024^3 particles in a cubic box, where each side has a length of $800 h^{-1}$ Mpc.

For CMASS-NGC galaxies with redshifts in the range of $z \in [0.45, 0.55]$, the spatial and redshift distributions are illustrated in Fig. 1. The COLA mock survey exhibits a similar angular distribution to that of CMASS-NGC. Additionally, the histograms of the number of galaxies for COLA and CMASS-NGC are displayed at the bottom panel. As shown, the redshift distribution for the simulations and observation are very similar, and the changes in the number of COLA mocks by varying Ω_m from 0.31 to 0.4 are small.

C. PATCHY simulation for covariance estimation

To accurately estimate the covariance matrix for MCFs, 1000 PATCHY catalogs [57] are employed in this study. The PATCHY (perturbation theory catalog generator of Halo and galaxy distributions) mock employs a method that integrates an efficient structure formation model with a local, nonlinear, scale-dependent, and stochastic biasing scheme to produce mock halo catalogs. Augmented Lagrangian perturbation theory is utilized for simulating the structure formation, combining 2LPT on large scales with the spherical collapse model on smaller scales. This technique generates a dark matter density field on a mesh starting from Gaussian fluctuations and calculates the peculiar velocity field. In short, PATCHY mock catalogs are generated using approximate gravity solvers and analytical-statistical biasing models, with halo occupation parameters adjusted such that the mocks well reproduce the BOSS two- and three-point statistics. They have been calibrated to the BigMultiDark N-body simulation with high resolution [58,59], which utilizes 3840^3 particles in a

¹<https://bitbucket.org/tassev/colacode/src/hg/>.

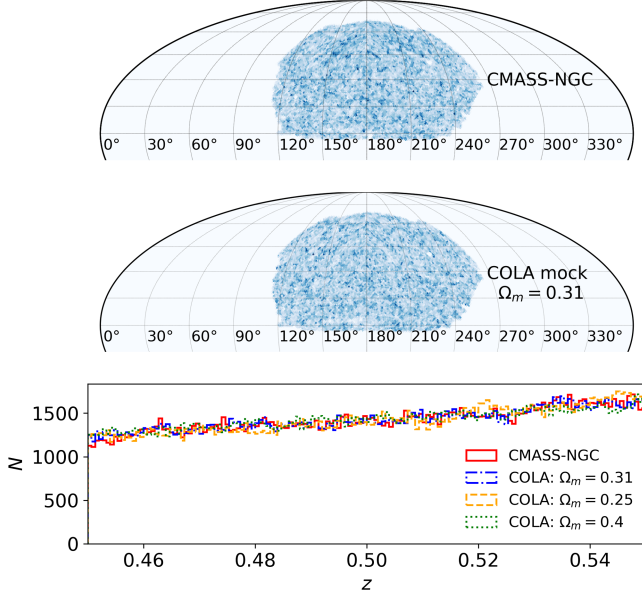


FIG. 1. Comparison of the spatial and redshift distributions for observational galaxies and simulation samples. Upper: the R.A.–Dec. distribution of CMASS-NGC galaxies in redshift space is illustrated for the range of $z \in [0.45, 0.55]$. Middle: the R.A.–Dec. distribution of generated COLA mock halos for $\Omega_m = 0.31$, showing the same angular distribution as CMASS-NGC galaxies. Bottom: the histograms of the number of galaxies (divided into 150 z -bins) are presented for CMASS-NGC (red solid) and COLA simulations with $\Omega_m = 0.31$ (blue dashed-dotted), 0.25 (orange dashed), and 0.4 (green dotted), respectively. As seen, the redshift distributions of the three COLA samples closely match that of CMASS-NGC.

volume of $(2.5 h^{-1} \text{Gpc})^3$. This simulation assumes a Λ CDM cosmology with $\Omega_m = 0.307$, $\Omega_b = 0.048$, $\sigma_8 = 0.82$, $n_s = 0.96$, and $h = 0.67$.

III. METHODOLOGY

A. β -skeleton analysis

We will perform the β -skeleton analysis and use its properties as weights to conduct the MCFs for CMASS-NGC data and COLA mocks. Let us first briefly summarize the β -skeleton theory and explain how it is used to characterize the cosmic web.

Stemming from the fields of computational geometry and geometric graph theory, the β -skeleton algorithm has been widely applied in image analysis, machine learning, visual perception, and pattern recognition [60–62]. In the realm of web finders, a class of algorithms, including the β -skeleton, is employed to construct a graph describing the degree of connectedness, starting from a set of 3D spatial points. This process, resembling the minimum spanning tree (MST) algorithm [63], has a key distinction—the resulting graph depends on the continuous β parameter. Additionally, web finders designed based on topological persistence, such as DisPerSE [64], Betti numbers [65], the

caustic skeleton defined based on Lagrangian fluid dynamics [66], and others, are related to the β -skeleton. For more detailed information regarding the β -skeleton in topology and geometric graph theory, please refer to [67,68].

In an n -dimensional Euclidean space, the edge set is defined by the β skeleton for a point set S . In this framework, points p and q in S are considered connected if there is no third point in the various empty regions. Specifically, for $0 < \beta < 1$, the empty region is formed by the intersection of all spheres with diameter d_{pq}/β , with p and q on their boundary. When $\beta = 1$, the empty region becomes the sphere with diameter d_{pq} . For $\beta \geq 1$, the empty region is defined in two different ways: the circle-based definition and the lune-based definition, the latter of which is adopted in this study. In this paper, according to this definition, the empty region R_{pq} is the intersection of two spheres with diameter βd_{pq} , centered at $p + \beta(q - p)/2$ and $q + \beta(p - q)/2$, respectively.

The β -skeleton, as defined above, possesses various interesting mathematical properties. As β continuously varies from 0 to ∞ , the constructed graphs transition from a complete graph to an empty graph. The special case of $\beta = 1$ results in the Gabriel graph, known to include the Euclidean MST. In image analysis, for instance, it has been utilized to reconstruct the shape of a two-dimensional object based on a set of sample points on the object’s boundary. It has been proven that the circle-based graphs with $\beta = 1.7$ can accurately reconstruct smooth surface boundaries without generating extra edges when samples are dense enough relative to local curvature. It is apparent that the volume of the excluded region exhibits a monotonically increasing relationship with β . As the excluded region expands, the requirements become more stringent and challenging, making it less likely to find connected paired galaxies. The β -skeleton code used in this paper is available at the following link.²

The β -skeleton algorithm is applied with β values of 1, 3, and 5, respectively, to both CMASS-NGC galaxies and COLA mock samples for various values of Ω_m . The derived cosmic webs are depicted in Fig. 2. For comparison, the Friend-of-Friend (FoF) algorithm is applied with linking lengths (denoted as L_{link}) set to 10 and 15 h^{-1} Mpc, and the resulting FoF cosmic webs are also displayed (bottom). All maps are selected within the range of R.A. $\in [200^\circ, 208^\circ]$ and Dec $\in [30^\circ, 38^\circ]$, corresponding to a slice of side length $183.96 \times 183.96 h^{-1}$ Mpc, respectively. The depth of these maps is $227.54 h^{-1}$ Mpc, corresponding to the redshift range of $z \in [0.45, 0.55]$.

Clearly, the skeletons of maps are observed through connections, and the number of connections decreases when increasing β , confirming the definition proposed

²<https://github.com/xiaodongli1986/LSSCode>.

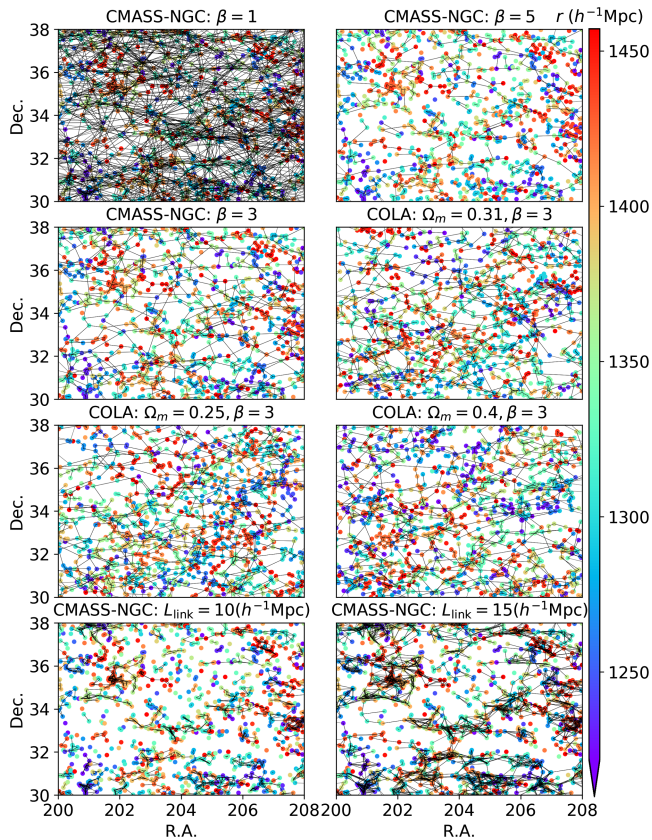


FIG. 2. Visualization of cosmic webs produced by the β -skeleton and Friend-of-Friend (FoF) algorithms for comparison. All maps are selected within the range of R.A. $\in [200^\circ, 208^\circ]$ and Dec $\in [30^\circ, 38^\circ]$. In the visualization, the scatter represents galaxies, and the resulting connections between galaxy pairs are also depicted. The color is indicative of the comoving distance of the respective galaxies. Upper: the β -skeleton web of CMASS-NGC galaxies is illustrated for $\beta = 1$ (left) and $\beta = 5$ (right). The second and third rows: the β -skeleton web is obtained by choosing $\beta = 3$ for CMASS-NGC and COLA simulations with $\Omega_m = 0.31$, $\Omega_m = 0.25$, $\Omega_m = 0.4$. Bottom: the cosmic web obtained from FoF with link lengths, L_{link} , set to 10 (left) and 15 (right), respectively. In comparison to $L_{\text{link}} = 10$ and 15 in FoF, the β -skeleton is adaptive to both low and high-density regions. As observed, the β -skeleton graph is more reliable and natural compared to the cosmic webs produced by FoF, highlighting the advantages of the β -skeleton graph. $\beta = 3$ is found to be optimal and better corresponds to the real cosmic web, represented as the plot of $\Omega_m = 0.31$, $\beta = 3$, in contrast to the two graphs that exhibit overly dense ($\beta = 1$) or overly sparse ($\beta = 5$) patterns.

for the β skeleton. Specifically, as β increases, the empty regions also increase. In essence, the requirement for establishing a connection between two galaxies becomes more stringent. Specifically, considering the CMASS-NGC plot, for $\beta = 3$, we detected 1182 connections, which is very comparable to the number of galaxies ($N_{\text{gal}} = 1345$) in each CMASS-NGC plot. However, at $\beta = 1$, we identified 3610 connections, significantly exceeding the

number of galaxy points. Conversely, for $\beta = 5$, we only detected 800 connections.

Moreover, filamentlike structures are generated from the galaxy sample by the β -skeleton; this is most clearly detected when $\beta = 3$. More densely and sparsely distributed filaments are produced by larger and smaller values of β , unlike cosmic webs observed in reality. Additionally, “knots” within the structure are formed by some of these galaxies, where three or more galaxies are connected. Moreover, there are isolated structures with a relatively small number of group members. Varying β strongly influences the overall shape of the skeleton maps. For instance, when $\beta = 1$, it roughly corresponds to computing two-point correlations, resulting in numerous connections being generated, irrespective of whether these connections lie within a filament or not. With $\beta = 3$, the generated set of structures closely approximates the observed cosmic web, with the number of connections comparable to the actual number of galaxies. In contrast, when $\beta = 5$, connections are sparse, as expected, identifying and connecting only the small and relatively isolated compact groups of galaxies. Moreover, statistically, the connection length, denoted as L_{con} , varies with β . As β increases, L_{con} tends to decrease and appears more concentrated. This is because the stringent threshold makes connecting two points separated by a large distance more challenging. Also, upon visual inspection, it may be challenging to identify obvious differences in the skeleton graphs corresponding to different values of Ω_m .

From these β -skeleton cosmic webs, we can quantify several relevant statistical measures. These include the comoving distance (r), the connection length of each pair of galaxies (L_{con}), the orientation of these connections (μ), and the number of surrounding neighbors of each galaxy (N_{con}).

In Fig. 3, we present the measured statistics of the resulting cosmic webs obtained by applying the β -skeleton algorithm with $\beta = 3$. In the left panel, the distribution of the comoving distance r of the nodes is shown. The red solid line represents the CMASS-NGC data, and for comparison, three different cases in the COLA simulations are considered with Ω_m values of 0.31 (blue dashed-dotted), 0.25 (orange-dashed), and 0.4 (green-dotted).

The deviations between the real data and COLA simulations, regardless of the Ω_m values used, are considerably small, with no significant differences observed. The mean comoving distance r is $1346.1 h^{-1} \text{ Mpc}$ for CMASS-NGC, $1345.7 h^{-1} \text{ Mpc}$ for $\Omega_m = 0.25$, $1346.7 h^{-1} \text{ Mpc}$ for $\Omega_m = 0.31$, and $1344.8 h^{-1} \text{ Mpc}$ for $\Omega_m = 0.4$. These values suggest that when varying Ω_m , the changes in r for COLA are on the order of 0.1%, making it challenging to rely solely on this quantity for distinguishing between cosmologies.

Similarly, the density distributions of the connection length L_{con} and the orientation of these connections, μ , are

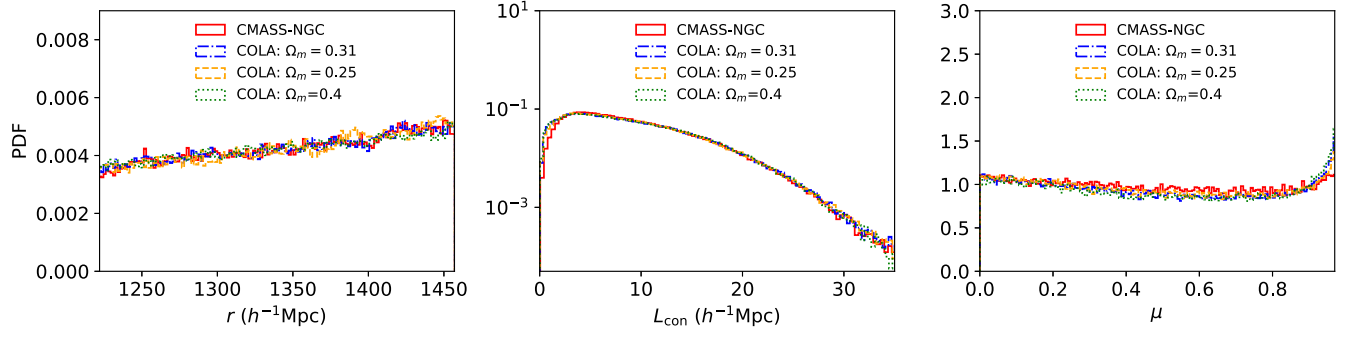


FIG. 3. Measured statistics of the β -skeleton cosmic webs for $\beta = 3$ in both the CMASS-NGC and COLA simulations, considering Ω_m values of 0.31, 0.25, and 0.4, respectively. Left: the normalized histogram of the comoving distance r of the nodes. Middle: the distribution of the connection length L_{con} . Right: the density distribution of the orientation of these connections, defined as $\mu = \cos(\theta)$, where θ is the angle between LOS and the connection line. As seen, the variation in these statistics concerning Ω_m is minimal, resulting in the β -skeleton cosmic web being insensitive solely to this cosmological parameter.

illustrated in the middle and right panels. We observe that the discrepancies in either L_{con} or μ between real data and COLA simulations are statistically insignificant. For instance, the mean L_{con} is $8.83 h^{-1} \text{ Mpc}$ for CMASS-NGC, and 8.70 , 8.68 , and $8.71 h^{-1} \text{ Mpc}$ for $\Omega_m = 0.25$, 0.31 , and 0.4 , respectively, leading to deviations of less than 2% across all cases. Additionally, the mean values of μ , averaged over the range of $\mu \in [0, 0.97]$, are 0.50 for CMASS-NGC, and 0.51, 0.52, 0.53 for $\Omega_m = 0.25$, 0.31 , 0.4 in COLA, respectively. This suggests that the directions of the connections in all cases are consistently randomly distributed, with no preferred orientation. Moreover, we can clearly observe a rise in density when $\mu > 0.9$. This is likely due to the Finger of God (FoG) effect, increasing the orientation along LOS.

In Fig. 4, we present a histogram illustrating the connectivity (denoted as N_{con}), representing the number

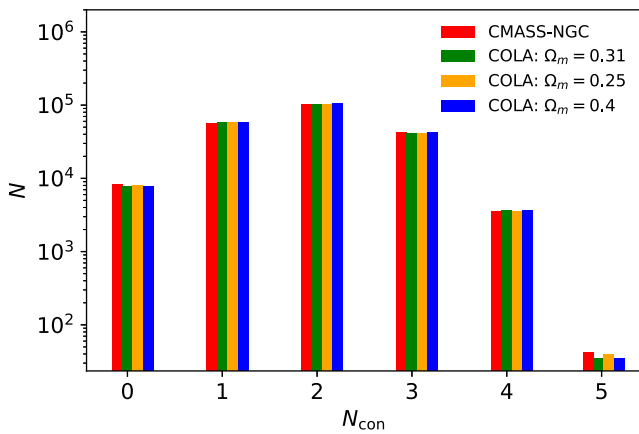


FIG. 4. Histogram of the number of neighbors around each galaxy, referred to as connectivity (N_{con}), in the range of $[0, 5]$ within the β -skeleton cosmic webs. It peaks at $N_{\text{con}} = 2$ and then decreases rapidly. The variation in connectivity is negligible for both CMASS-NGC data and COLA mocks under three different Ω_m values.

of neighbors around each galaxy, displaying the values within the range of $[0, 5]$ for the obtained β -skeleton cosmic webs. We observe 8284, 57111, 102639, 41781, 3581 galaxies with $N_{\text{con}} = 0, 1, 2, 3, 4$, respectively, along with only 42 galaxies having $N_{\text{con}} = 5$. In the case of COLA galaxies under $\Omega_m = 0.31$, the numbers of galaxies are 7857, 57261, 103680, 41642, 3657, and 35, respectively, for N_{con} ranging from 0 to 5. Notably, the variation in connectivity is found to be negligible across three distinct Ω_m values for the COLA mocks. Therefore, the sensitivity of N_{con} to Ω_m is insignificant.

In Fig. 5, we illustrate the averaged connection lengths (denoted as \bar{D}_{nei}) between each galaxy and its surrounding neighbors, where neighbor galaxies are determined by their skeleton graph in the case of $\beta = 3$. On the left side, the histograms (divided into 150 bins) display the number of galaxies over \bar{D}_{nei} for both CMASS-NGC data and COLA mocks across $\Omega_m = 0.31, 0.25$, and 0.4 , respectively. In all cases, the histograms peak at $\bar{D}_{\text{nei}} = 0$, indicating that galaxies with no connections dominate the cosmic webs.

We observe that within the range of \bar{D}_{nei} from 0 to $30 h^{-1} \text{ Mpc}$ for CMASS-NGC, the total number of galaxies is 213,438. However, beyond this range, the total number decreases exponentially to 86. For the COLA mocks, we find that for $\Omega_m = 0.31$, the total number is 214,132 within the range of \bar{D}_{nei} from 0 to $30 h^{-1} \text{ Mpc}$, and 95 beyond this range. For $\Omega_m = 0.25$ and 0.4 , the corresponding totals are 213,428 and 214,472 within the interval of $[0, 30] h^{-1} \text{ Mpc}$, and 103 and 69 outside this range. Moreover, the mean value of \bar{D}_{nei} is approximately $8 \text{ Mpc}/h$ across all cases.

On the right side, four violin plots depict the dependence of \bar{D}_{nei} distributions on connectivity N_{con} , ranging from 1 to 5. The cases of CMASS-NGC data and COLA mocks under three different Ω_m values are presented sequentially. As seen, all COLA results are nearly identical to the CMASS-NGC case when $N_{\text{con}} < 5$. However, when $N_{\text{con}} = 5$, the distribution of \bar{D}_{nei} exhibits slight variation, with mean values of

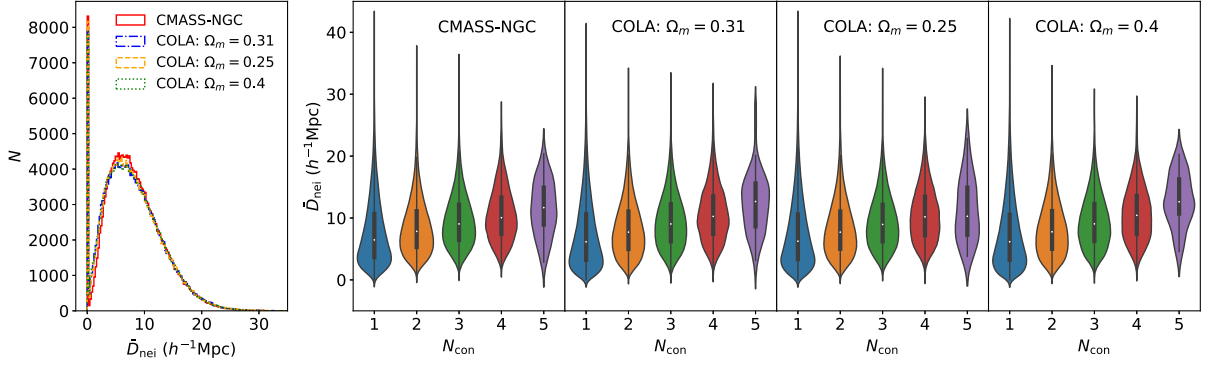


FIG. 5. Distribution of the averaged connection length between each galaxy and its surrounding neighbor galaxies, denoted as \bar{D}_{nei} , in the case of $\beta = 3$, where the neighbor galaxies are determined by their skeleton graph. Left: the histograms display the number of galaxies over \bar{D}_{nei} for both CMASS-NGC data and COLA mocks under three distinct Ω_m values. The first peak is located at $\bar{D}_{\text{nei}} = 0$, indicating that galaxies with no connections dominate the cosmic webs for $\beta = 3$. The mean value of \bar{D}_{nei} is approximately $8 h^{-1} \text{Mpc}$ among all cases. Right (four violin plots): dependence of the distributions of \bar{D}_{nei} on N_{con} , ranging from 1 to 5. We present, in sequence, the cases of CMASS-NGC data and COLA mocks for $\Omega_m = 0.31, 0.25, 0.4$, respectively.

$\bar{D}_{\text{nei}} = 11.52, 12.38, 11.27, 12.78 h^{-1} \text{Mpc}$ for CMASS-NGC and COLA simulations with $\Omega_m = 0.31, 0.25$, and 0.4 .

The above analysis is based on the statistics of the β -skeleton graphs. We observe that relying on these statistics alone does not enable us to distinguish between cosmologies with significantly different values of Ω_m . Consequently, in the following we will demonstrate that by utilizing the obtained β -skeleton statistics as weights, the MCF method exhibits high sensitivity to Ω_m .

B. MCFs using the β -skeleton statistics

Next, we will perform an analysis through MCFs using the weights from the β -skeleton statistics. A standard cosmological analysis generally involves the computation of the 2PCF to infer cosmological information from galaxy clustering properties. Meanwhile, MCFs assign weights to different features of galaxies to extract non-Gaussian information on LSS effectively capturing more detailed clustering information. Considering that the β -skeleton statistics provide different measures of LSS, we then use these statistics as weights in MCFs to study their sensitivities to the cosmological parameter, the present-day matter density Ω_m .

In the previous sections, we have introduced several statistical measures for β -skeleton graph. We will specifically focus on the following three statistics as the weights of MCFs: the number of neighbors around each galaxy, N_{con} , the averaged connection length between each galaxy and its surrounding neighbors, \bar{D}_{nei} , and its reciprocal, $1/\bar{D}_{\text{nei}}$. Note that, in the following, all MCF analyses will be performed with β fixed at 3 for the β -skeleton measurements. This is the first time to perform an analysis on real and mock LSS data by combining MCFs and the β -skeleton statistics.

Briefly, the procedure for calculating the MCFs remains the same as the standard one; however, the weights are assigned based on the measures of the β -skeleton statistics. Compared with the traditional 2PCF, defined as $\xi(\mathbf{r}) = \langle \delta(\mathbf{x})\delta(\mathbf{x} + \mathbf{r}) \rangle$, the form of MCF, following [45], is given by

$$W(\mathbf{r}) = \langle \delta(\mathbf{x})w(\mathbf{x})\delta(\mathbf{x} + \mathbf{r})w(\mathbf{x} + \mathbf{r}) \rangle, \quad (1)$$

where the term $w(\mathbf{x})$ denotes the chosen weights used in MCFs. The term $\delta(\mathbf{x})$ denotes the pointlike density contrast, expressed as $\delta(\mathbf{x}) = \delta\rho/\bar{\rho}$.

As mentioned earlier, three different weights are chosen for each galaxy in this study, specifically,

$$w(\mathbf{x}) = \begin{cases} N_{\text{con}}(\mathbf{x}), & \text{case I} \\ \bar{D}_{\text{nei}}(\mathbf{x}), & \text{case II} \\ 1/\bar{D}_{\text{nei}}(\mathbf{x}), & \text{case III} \end{cases}. \quad (2)$$

Here, $w(\mathbf{x})$ is obtained from the β -skeleton at the position \mathbf{x} . Note that the β -skeleton graph may have a fraction of samples (detected by knots in the skeleton) compared to the full observed samples of galaxies. Consequently, some galaxies may lack these statistical measurements. In such cases, we assign $w(\mathbf{x}) = 0$. Additionally, when $\bar{D}_{\text{nei}} = 0$, to prevent divergence for $1/\bar{D}_{\text{nei}}$, we also assign $w(\mathbf{x}) = 0$ for case III.

We utilize the widely adopted Landy–Szalay estimator [69], expressed as

$$W(s, \mu) = \frac{WW - 2WR + RR}{RR}, \quad (3)$$

Here, WW denotes the weighted count of galaxy-galaxy pairs, WR corresponds to galaxy-random pairs, and RR

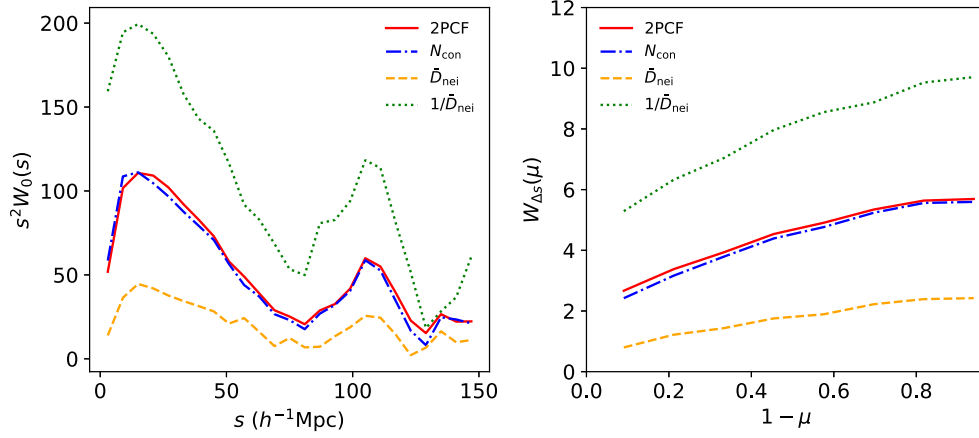


FIG. 6. Comparison of one-dimensional MCFs with different weighting schemes for the CMASS-NGC dataset. Left: In comparison to the 2PCF (red solid), the monopole of MCFs, $s^2 W_0(s)$ [as defined in Eq. (4)], is shown for three different weights: N_{con} (blue dashed-dotted), \bar{D}_{nei} (orange dashed), and $1/\bar{D}_{\text{nei}}$ (green dotted). Right: Same as in the left panel, but for focusing on the anisotropic clustering in MCFs, $W_{\Delta s}(\mu)$ [as defined in Eq. (5)]. As observed, using the weight $1/\bar{D}_{\text{nei}}$ leads to the highest magnitudes in both panels. This is because a number of objects in underdense regions are assigned a larger weight, significantly enhancing the amplitudes of these two-point functions. Conversely, assigning \bar{D}_{nei} as the weight will lead to the lowest amplitudes in both panels. In either the left or right panel, the MCF with N_{con} as the weight results in the amplitude and shape very comparable to those of the 2PCF, implying it is not statistically related to the local properties of each galaxy.

represents random-random pairs. These pairs are determined within a distance defined by $s \pm \Delta s$ and $\mu \pm \Delta\mu$, where s represents the distance between pairs, and μ is defined as $\cos(\theta)$. Here, θ is the angle between the line of sight (LOS) direction and the line connecting the pair.

The random sample comprises 10 times the number of objects compared to the CMASS-NGC galaxies, and each PATCHY simulation for covariance estimation includes 20 times the number of objects than CMASS-NGC. Based on our testing, results converge with a random sample containing about 10 times more objects than CMASS-NGC, and the weights for the random samples are consistently set to unity.

C. One-dimensional MCFs: $\hat{W}_0(s)$ and $\hat{W}_{\Delta s}(\mu)$

By integrating $W(s, \mu)$ over either s or μ , two one-dimensional statistical quantities can be defined. The first one is the monopole of the MCF, a function of the clustering scale, represented as

$$W_0(s) = \int_0^1 W(s, \mu) d\mu. \quad (4)$$

The second quantity is the μ -dependent function, represented as

$$W_{\Delta s}(\mu) = \int_{s_{\min}}^{s_{\max}} W(s, \mu) ds. \quad (5)$$

The values $s_{\min} = 15 h^{-1}\text{Mpc}$ and $s_{\max} = 40 h^{-1}\text{Mpc}$ have been used for quantifying both redshift-space distortions

(RSDs) and Alcock-Paczyński (AP) distortions within the context of the tomographic AP method [70].

To mitigate the impact of galaxy bias and enhance the accuracy of the analysis, it is common to employ normalized statistics, which utilize the shape rather than the amplitude to extract cosmological information. By doing this, the normalized quantities based on Eqs. (4) and (5) can be expressed as follows:

$$\begin{aligned} \hat{W}_0(s) &= \frac{W_0(s)}{\int_a^b W_0(s) ds}, \\ \hat{W}_{\Delta s}(\mu) &= \frac{W_{\Delta s}(\mu)}{\int_0^{\mu_{\max}} W_{\Delta s}(\mu) d\mu}. \end{aligned} \quad (6)$$

In choosing $a = 15 h^{-1}\text{Mpc}$, $b = 57 h^{-1}\text{Mpc}$, and $\mu_{\max} = 0.97$ (to reduce effects of fiber collisions and RSDs), extensive testing indicates that the selected parameters are effective for studying clustering features and for enhancing the sensitivity on different cosmologies. From tests, we find that opting for equally spaced divisions of s within the range of $s \in [15, 57] h^{-1}\text{Mpc}$ and μ in the interval $\mu \in [0, 0.97]$ into seven bins results in optimal and robust constraints on Ω_m .

IV. RESULTS

Figure 6 illustrates an analysis of the one-dimensional MCFs utilizing various weighting schemes on the CMASS-NGC dataset. On the left panel, contrasting with the 2PCF (red solid), the monopoles of MCFs, $s^2 W_0(s)$ [as defined in Eq. (4)], are illustrated for three distinct weights: N_{con} (blue dashed-dotted), \bar{D}_{nei} (orange dashed), and $1/\bar{D}_{\text{nei}}$ (green

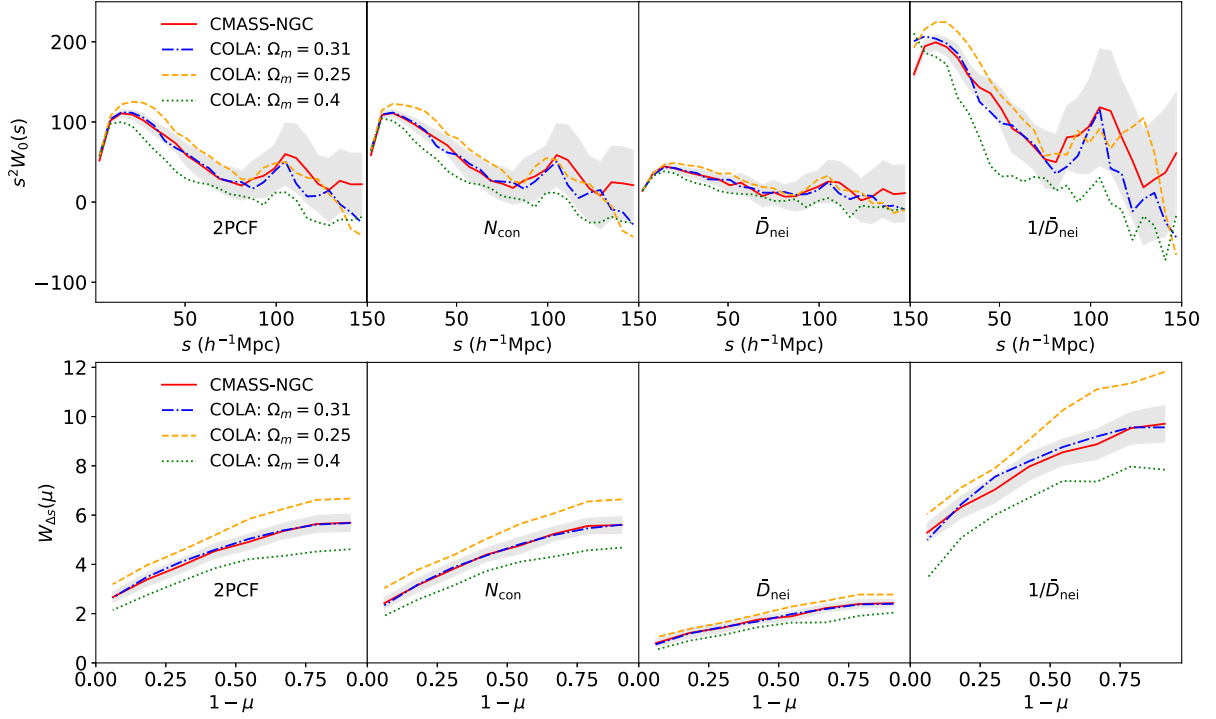


FIG. 7. Comparison of one-dimensional MCFs $s^2 W_0(s)$ (top) and $W_{\Delta s}(\mu)$ (bottom) for four datasets: CMASS-NGC (red solid), COLA mocks with $\Omega_m = 0.31$ (blue dashed-dotted), $\Omega_m = 0.25$ (orange dashed), $\Omega_m = 0.4$ (green dotted), respectively. From left to right, we present the results for the 2PCF and MCFs using the weights N_{con} , \bar{D}_{nei} , and $1/\bar{D}_{\text{nei}}$, respectively. The gray-shaded region represents 2σ errors estimated using PATCHY mocks. For the COLA mock with $\Omega_m = 0.31$, all derived MCFs in the s and μ regions we considered ($s \in [0, 150] h^{-1} \text{Mpc}$, $\mu \in [0, 0.97]$) match well with those of CMASS-NGC within 2σ level. However, when $\Omega_m = 0.4$ or 0.25 in COLA largely deviates from the true value in CMASS-NGC, both 2PCFs and MCFs become significantly inconsistent with the observational data beyond 2σ level.

dotted). The right panel depicts the same comparison but emphasizes the anisotropic clustering in MCFs, $W_{\Delta s}(\mu)$ [as defined in Eq. (5)]. Notably, the use of the weight $1/\bar{D}_{\text{nei}}$ yields the highest magnitudes in both panels. This is attributed to the assignment of a larger weight to numerous objects in underdense regions, significantly enhancing the amplitudes of these MCFs. Conversely, assigning \bar{D}_{nei} as the weight results in the lowest amplitudes. Furthermore, in either panel, employing N_{con} as the weighting scheme for MCFs results in the amplitude and the shape that agree with those of the 2PCF. This is because N_{con} is not statistically related to the clustering strength of the structure (e.g., a galaxy in a low-density region can have a larger N_{con} than a galaxy in a high-density region). Therefore, using it as weights does not significantly alter the statistical clustering.

In Fig. 7, a comparison of one-dimensional MCFs $s^2 W_0(s)$ (top) and $W_{\Delta s}(\mu)$ (bottom) is presented for four datasets: CMASS-NGC (red solid), COLA mocks with $\Omega_m = 0.31$ (blue dashed-dotted), $\Omega_m = 0.25$ (orange dashed), and $\Omega_m = 0.4$ (green dotted), respectively. The results for the 2PCF and MCFs using the weights N_{con} , \bar{D}_{nei} , and $1/\bar{D}_{\text{nei}}$ are displayed from left to right. The gray-shaded region represents 2σ errors estimated using PATCHY mocks. For the COLA mock with $\Omega_m = 0.31$,

all derived MCFs in the s and μ regions ($s \in [0, 150]$, $\mu \in [0, 0.97]$) match well with those of CMASS-NGC within the 2σ level. However, when $\Omega_m = 0.4$ or 0.25 in COLA, there is a significant deviation from the true value in CMASS-NGC, causing both 2PCFs and MCFs to become markedly inconsistent with the observational data beyond the 2σ level.

Moreover, Fig. 8 illustrates the normalized MCFs, $s^2 \hat{W}_0(s)$ (top) within the range $s \in [15, 57] h^{-1} \text{Mpc}$, and $\hat{W}_{\Delta s}(\mu)$ (bottom) within the range $\mu \in [0, 0.97]$, evenly spaced and divided into seven bins. The gray-shaded region denotes 2σ errors estimated from PATCHY mocks. The results of COLA mocks under $\Omega_m = 0.31$ are well consistent with those of CMASS-NGC. However, deviations become noticeable for larger and smaller Ω_m values, particularly for the statistic $s^2 \hat{W}_0(s)$. Significantly, the weight $1/\bar{D}_{\text{nei}}$ serves as a sensitive indicator for Ω_m , where an incorrect value can lead to substantial deviations from those observed in the CMASS-NGC data, exceeding the 2σ level.

To test the potential improvement of cosmological constraints using β -skeleton weighting schemes, we select $\hat{W}_0(s)$ and $\hat{W}_{\Delta s}(\mu)$ measurements. We utilize the χ^2 statistic to quantitatively distinguish the wrong cosmologies from the correct one and evaluate the performance in

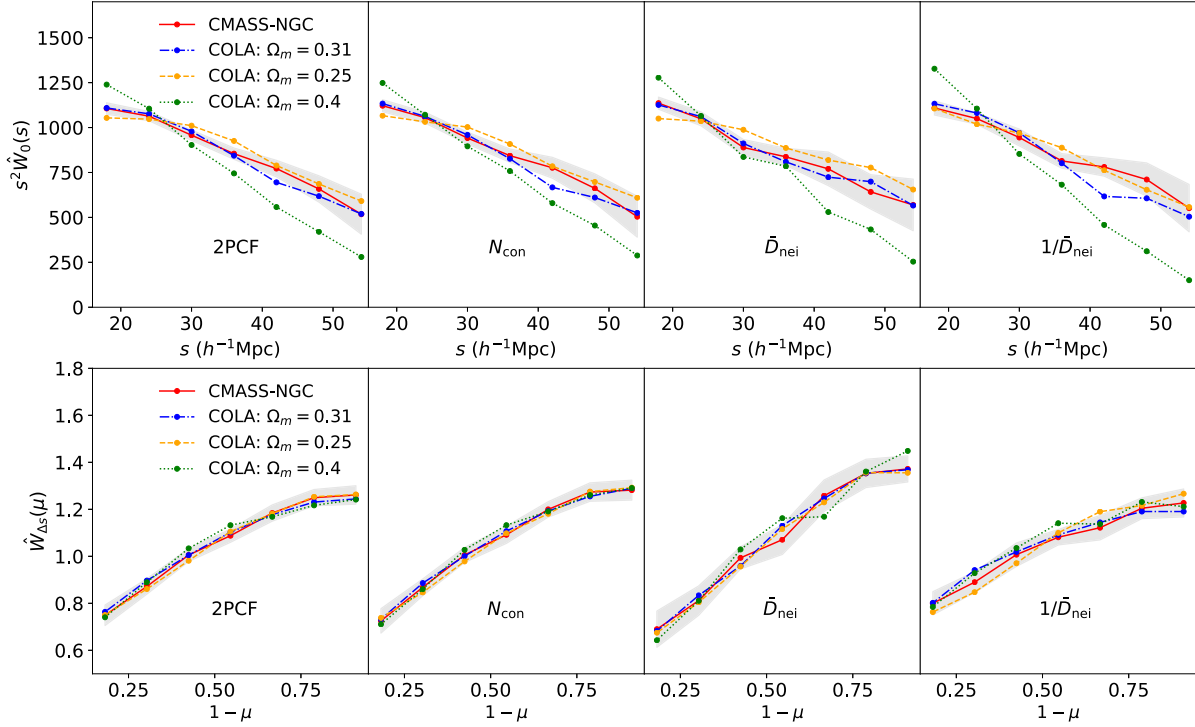


FIG. 8. Similar to Fig. 7, but for the normalized MCFs, $s^2 \hat{W}_0(s)$ (top) in the range of $s \in [15, 57] h^{-1}$ Mpc, and $\hat{W}_{\Delta s}(\mu)$ (bottom) in the range of $\mu \in [0, 0.97]$, equally spaced and divided into seven bins. The gray-shaded region represents 2σ errors estimated by the PATCHY mocks. We observe that the results of COLA mocks under $\Omega_m = 0.31$ agree well with those of CMASS-NGC, whereas larger and smaller Ω_m values can lead to apparent deviations, especially for the statistic $s^2 \hat{W}_0$. Additionally, we note that the weight $1/\bar{D}_{\text{nei}}$ can provide a sensitive probe to Ω_m , as an incorrect value can result in significant deviations from those of the CMASS-NGC data beyond the 2σ level.

constraining power. The χ^2 function for fitting the data is given by

$$\chi^2 = (\Delta \mathbf{p})^T \cdot \mathbf{C}^{-1} \cdot \Delta \mathbf{p}, \quad (7)$$

where

$$\Delta \mathbf{p} = \mathbf{p}_{\text{model}} - \mathbf{p}_{\text{data}}. \quad (8)$$

Here, $\mathbf{p}_{\text{model}} = \mathbf{p}_{\text{COLA}}(\Omega_m)$ represents various statistical quantities of the COLA mocks, including the 2PCF, $\hat{W}_0(s)$, $\hat{W}_{\Delta s}(\mu)$, and their combinations. The symbol \mathbf{p}_{data} denotes the measurements from the observational CMASS-NGC data. By varying Ω_m within the COLA simulations for $\Omega_m = 0.25$, $\Omega_m = 0.31$, and $\Omega_m = 0.4$, the resulting χ^2 values are then used to assess the sensitivity of the proposed statistics to Ω_m .

The corresponding covariance matrix, denoted by \mathbf{C} , is evaluated using the PATCHY simulation mocks. Specifically, the empirical covariance matrix of a vector \mathbf{p} is given by

$$\mathbf{C} = \langle (\mathbf{p} - \bar{\mathbf{p}})(\mathbf{p} - \bar{\mathbf{p}})^T \rangle \quad (9)$$

$$= \frac{1}{N_{\text{mock}} - 1} \sum_{i=1}^{N_{\text{mock}}} (\mathbf{p}_i - \bar{\mathbf{p}})(\mathbf{p}_i - \bar{\mathbf{p}})^T, \quad (10)$$

by averaging over all N_{mock} mock samples. Here, \mathbf{p}_i , with a length of N_p , denotes a vector containing all statistical quantities for the i th mock sample. Note that the mean of \mathbf{p} over all mocks is denoted by $\bar{\mathbf{p}}$. We utilized 1000 PATCHY mocks to accurately estimate the relevant covariance matrices for the various considered statistics.

In Fig. 9, the correlation coefficient matrices of the 2PCF, $\hat{W}_0(s)$ (top) and $\hat{W}_{\Delta s}(\mu)$ (bottom), along with their combinations, are illustrated. To clearly depict the correlation strength between different statistics, each matrix displays the cross and auto-correlation coefficients between 2PCF and MCFs for three weighting schemes: N_{con} , \bar{D}_{nei} , and $1/\bar{D}_{\text{nei}}$. The correlation between 2PCF and MCFs is notably strong for the weight of N_{con} , while such correlations between 2PCF and the other weights become relatively weaker. Additionally, the correlations for combinations of the total of four statistical measurements, including 2PCF and MCFs with the three weights, are shown. These combinations are used to estimate the χ^2 values for the joint analysis.

To quantitatively assess the sensitivity of various statistical measurements for different Ω_m values, we further define the change of χ^2 deviating from the same statistical measurement under the fiducial value of $\Omega_m = 0.31$ (the corresponding χ^2 is denoted by χ_{min}^2),

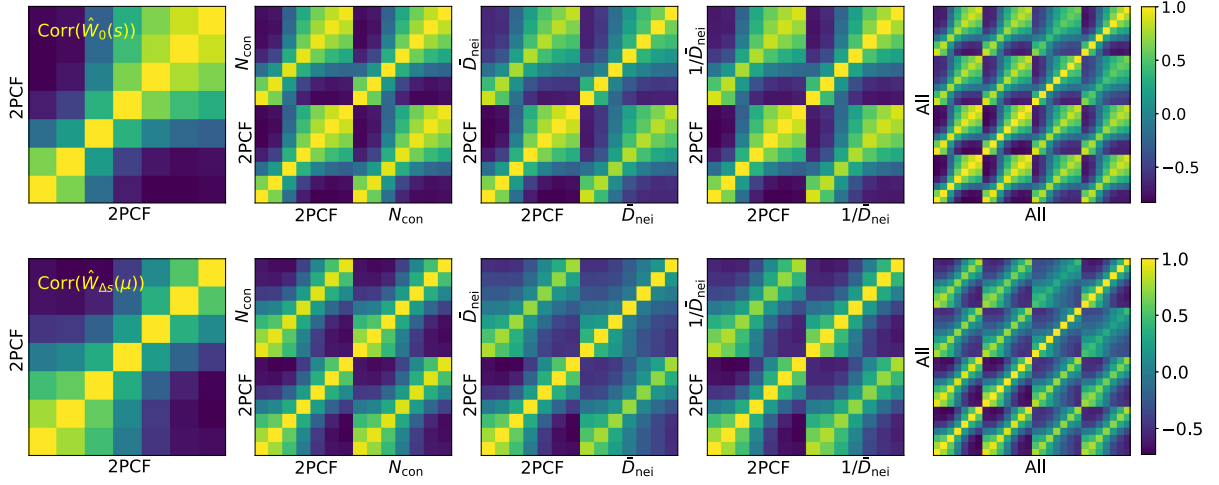


FIG. 9. Correlation coefficient matrices of 2PCF, $\hat{W}_0(s)$ (top) and $\hat{W}_{\Delta_s}(\mu)$ (bottom). The correlation coefficients are estimated using the 1000 PATCHY mocks, where $\Omega_m = 0.31$. To clearly illustrate the correlation strength between different statistics, each matrix displays the cross and auto-correlation coefficients between 2PCF and MCF with three weighting schemes for N_{con} , \bar{D}_{nei} , and $1/\bar{D}_{\text{nei}}$.

$$\Delta\chi^2 \equiv \chi^2 - \chi_{\text{min}}^2. \quad (11)$$

Displayed in Fig. 10 are the changes in χ^2 values, according to Eq. (11), for $\hat{W}_0(s)$ (left) and $\hat{W}_{\Delta_s}(\mu)$ (right) while Ω_m ranges from 0.25 to 0.4. The calculation of χ_{min}^2 utilizes the COLA mock with $\Omega_m = 0.31$ following Eq. (7), where the minima in χ^2 values are reached compared to those for the other two Ω_m values. As observed, MCFs exhibit a notably distinct dependence of χ^2 on Ω_m compared to the 2PCF. As expected, the use of a combination of 2PCF and MCFs with the β -skeleton weights yields the

highest sensitivity to Ω_m . This is evident from the rapid changes in χ^2 when varying Ω_m .

Moreover, the relative changes of $\Delta\chi^2$ with respect to that of the 2PCF, defined as

$$r = \frac{\Delta\chi^2}{\Delta\chi_{\text{2PCF}}^2} - 1. \quad (12)$$

The r values are summarized in Table I. These relative changes can offer another indication of the sensitivity to Ω_m . The measurements involve various combinations of 2PCF and MCFs with three different weights. The

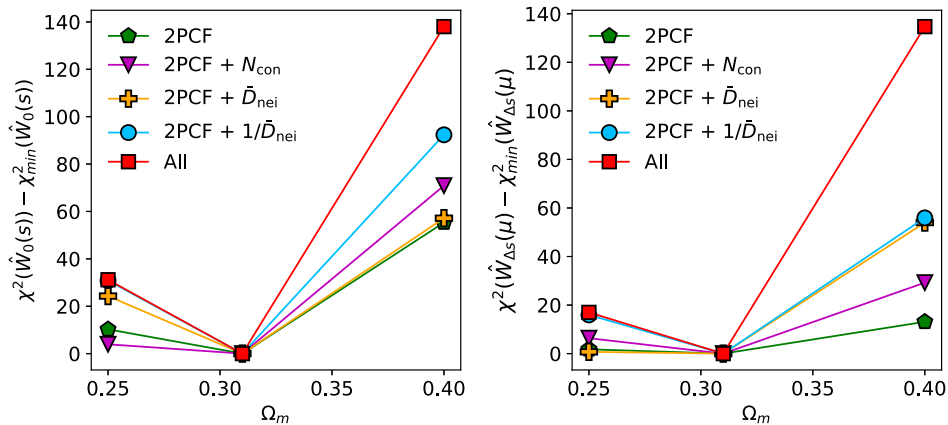


FIG. 10. Changes of χ^2 values [as defined in Eq. (7)] for $\hat{W}_0(s)$ (left) and $\hat{W}_{\Delta_s}(\mu)$ (right) as Ω_m varies from 0.25 to 0.4. Here, χ_{min}^2 is calculated by using $\Omega_m = 0.31$ for χ^2 , where the χ^2 values reach the minimum for either left and right panel compared to those for the other two Ω_m values. As seen, compared to the 2PCF, MCFs provide an apparently different dependence of χ^2 on Ω_m . As expected, employing a combination of 2PCF and MCFs with the three β -skeleton weights results in the highest sensitivity to Ω_m . This is evident from the rapid changes in χ^2 when Ω_m is varied.

TABLE I. Comparison of relative changes of χ^2 with respect to that of 2PCF, i.e., the r value as defined in Eq. (12), for different measurements with varying Ω_m from 0.25 to 0.4. The measurements are based on various combinations of 2PCF and MCFs with three different weights. The overall joint analysis, containing all four statistics (labeled as “All (I)” for the combination of 2PCF and three \hat{W}_0 statistics and “All (II)” for 2PCF and three \hat{W}_{Δ_s} statistics), shows the highest sensitivity to Ω_m in either the \hat{W}_0 or \hat{W}_{Δ_s} measurements. Such relative changes indicate that the combination of the 2PCF and three MCFs leads to a significant improvement in Ω_m constraint. For instance, the χ^2 values for $\Omega_m = 0.25$ are enhanced by 204.06% for “All (I)” and 855.47% for “All (II)”. Similarly, combining these two yields enhancements of 149.49% and 928.33%, respectively.

Statistics	$\Omega_m = 0.25$	$\Omega_m = 0.40$
2PCF	0	0
2PCF + $\hat{W}_0(N_{\text{con}})$	-61.61%	28.14%
2PCF + $\hat{W}_0(\bar{D}_{\text{nei}})$	136.97%	3.33%
2PCF + $\hat{W}_0(1/\bar{D}_{\text{nei}})$	200.91%	66.94%
All (I)	204.06%	149.49%
2PCF	0	0
2PCF + $\hat{W}_{\Delta_s}(N_{\text{con}})$	257.52%	123.15%
2PCF + $\hat{W}_{\Delta_s}(\bar{D}_{\text{nei}})$	-58.80%	312.19%
2PCF + $\hat{W}_{\Delta_s}(1/\bar{D}_{\text{nei}})$	797.33%	326.86%
All (II)	855.47%	928.33%

comprehensive joint analysis, including all four statistics [labeled as “All (I)” for the combination of 2PCF and three \hat{W}_0 statistics, and “All (II)” for 2PCF and three \hat{W}_{Δ_s} statistics], exhibits the highest sensitivity to Ω_m in either the \hat{W}_0 or \hat{W}_{Δ_s} measurements. These relative changes indicate a substantial enhancement in the constraint on Ω_m through the combination of 2PCF and three MCFs. For instance, the χ^2 values for $\Omega_m = 0.25$ are enhanced by 204.06% for “All (I)” and 855.47% for “All (II)”. Likewise, when these statistics are combined, enhancements of 149.49% and 928.33% are achieved for $\Omega_m = 0.40$, respectively.

V. CONCLUDING REMARKS

The β -skeleton approach proves to be a convenient tool for constructing the cosmic web based on the spatial geometry distribution of galaxies, especially in sparse samples. It plays a crucial role in establishing the three-dimensional structure of the Universe and serves as a quantitative tool for characterizing the nature of the cosmic web. Moreover, MCFs, by assigning weights to different features of galaxies to extract non-Gaussian information on large-scale structure (LSS), have proven effective in capturing more detailed clustering information.

This study marks the first application of combining mark weighted correlation function (MCFs) and β -skeleton

statistics to real observational data. We have introduced the β -skeleton information as weights in MCFs, presenting a novel statistical measure. The study applies the β -skeleton approach to the CMASS NGC galaxy samples from SDSS BOSS DR12 in the redshift interval $0.45 \leq z \leq 0.55$. Additionally, to evaluate the measurements for different cosmologies, we conducted three COLA cosmological simulations with different settings ($\Omega_m = 0.25$, $\Omega_m = 0.31$, $\Omega_m = 0.4$) for comparison.

We have measured three MCFs, each weighted by (i) the number of neighboring galaxies around each galaxy, N_{con} ; (ii) the average distance of each galaxy from its surrounding neighbors, \bar{D}_{nei} ; and (iii) the reciprocal of the average distance of each galaxy from its surrounding neighbors, $1/\bar{D}_{\text{nei}}$. Through a comparison of measurements and the calculation of corresponding χ^2 statistics, we observe a substantial improvement in the constraints on the cosmological parameter Ω_m by conducting a joint analysis of the standard 2PCF and all three MCFs for different weights.

In the joint analysis, the highest sensitivity to Ω_m is observed in either the \hat{W}_0 (monopoles of MCFs) or \hat{W}_{Δ_s} (anisotropic clustering of MCFs) measurements. The χ^2 values for the joint analysis are improved by approximately 150%–928% compared to the 2PCF alone. Overall, the joint analysis robustly enhances sensitivity to Ω_m , allowing for the rejection of incorrect cosmologies at a significant level.

Our study has introduced a novel MCF weighting strategy using β -skeleton information to maximize the extraction of LSS information. This approach holds the potential to be extended to other surveys and datasets, contributing to the constraint of cosmological parameters. In future work, simulation mocks for a wide range of cosmological parameters can be generated through emulator approaches, allowing for a thorough assessment of parameter constraints. The method presented here is expected to serve as a valuable analysis tool for upcoming stage-IV surveys, including the Chinese Space Station Telescope (CSST).

ACKNOWLEDGMENTS

This work is supported by National SKA Program of China (Grants No. 2020SKA0110401, No. 2020SKA0110402, No. 2020SKA0110100), the National Key R and D Program of China (Grants No. 2020YFC2201600, No. 2018YFA0404504, No. 2018YFA0404601), the National Science Foundation of China (Grants No. 11890691, No. 12203107, No. 12073088, No. 12373005), the China Manned Space Project with No. CMS-CSST-2021 (Projects No. A02, No. A03, No. B01), the Guangdong Basic and Applied Basic Research Foundation (Grant No. 2019A151511098), and the 111 project of the Ministry of Education No. B20019. We

also wish to acknowledge the Beijing Super Cloud Center (BSCC) and Beijing Beilong Super Cloud Computing Co., Ltd for providing HPC resources that have significantly contributed to the research results presented in this paper.

APPENDIX

To ensure that the accuracy of COLA simulation is sufficient for this analysis, here we have made a comparison between the statistics measured in COLA and GADGET [71–73] simulations. GADGET is a massively parallel TreeSPH code for accurate N-body/SPH simulations [74], which can simulate a collisionless fluid by the N-body method or an ideal gas with smoothed particle hydrodynamics (SPH). [50] combines the subhalo abundance matching (SHAM) procedure with COLA and GADGET simulations to create mocks for BOSS CMASS NGC galaxies by fitting the correlation function at $s \in [4, 20] h^{-1} \text{Mpc}$ within the redshift interval $z \in [0.45, 0.55]$. Below we compare the clustering properties of the COLA, GADGET, and PATCHY mocks with the observed data.

To ensure the reliability of our analysis, the COLA and GADGET mocks used the same cosmological parameters $\{\Omega_m = 0.2951, \Omega_\Lambda = 0.7049, \Omega_b = 0.0468, w = -1.0, \sigma_8 = 0.80, n_s = 0.96\}$ for simulating DM particles and ran simulations under identical conditions with 1024^3 particles in a box measuring $800 h^{-1} \text{Mpc}$.

In Fig. 11, we present a comparison of the 2PCF $s^2 W_0(s)$ (top) obtained from four datasets: CMASS-NGC (red solid), COLA mock with $\Omega_m = 0.2951$ (blue dashed-dotted), GADGET with $\Omega_m = 0.2951$ (orange dashed), and PATCHY with $\Omega_m = 0.307$ (green dotted). We also display their ratios to that of CMASS-NGC, W_0/W_0^{CMASS} (bottom). The gray shaded region represents the 2σ error estimated from the PATCHY mocks.

Observably, utilizing the SHAM procedure by fitting the 2PCF in the range of $s \in [4, 20] h^{-1} \text{Mpc}$ to that of CMASS-NGC, both the 2PCFs of COLA and GADGET agree with the data on the fitting scale within the 2σ level. The COLA and GADGET simulations do not share the same initial condition. Therefore, they have differences caused by the cosmic variance. Due to the use of a different Ω_m in the PATCHY simulation, 2PCF of PATCHY is essentially on the edge of the 2σ uncertainty. Furthermore, at smaller scales ($s < 4 h^{-1} \text{Mpc}$), we observe deviations for COLA and GADGET from the data slightly exceeding the 2σ level. This discrepancy arises because this scale extends beyond our fitting range in the SHAM procedure.

To further validate the consistency of the COLA mock, we conducted a comparison of the normalized 2PCF and MCFs derived from the CMASS-NGC data, GADGET, and PATCHY. This comparison focuses on larger clustering scales within the range of $s \in [15, 57] h^{-1} \text{Mpc}$, which

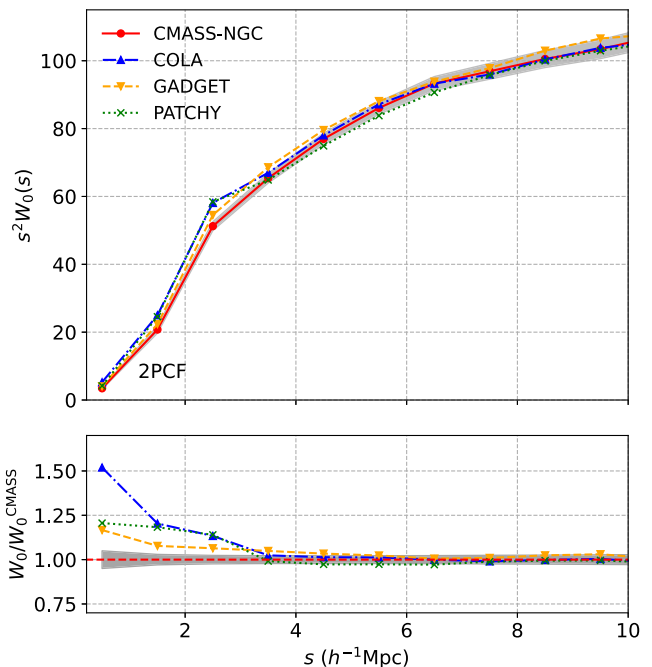


FIG. 11. Comparison of 2PCFs $s^2 W_0(s)$ for CMASS-NGC (red solid), COLA mock with $\Omega_m = 0.2951$ (blue dashed-dotted), GADGET with $\Omega_m = 0.2951$ (orange dashed), and PATCHY with $\Omega_m = 0.307$ (green dotted), along with their ratios to CMASS-NGC, W_0/W_0^{CMASS} (bottom). The gray shaded region indicates the 2σ error estimated from the PATCHY mocks. 2PCFs of COLA and GADGET agree with the data within 2σ on the fitting scale ($s \in [4, 20] h^{-1} \text{Mpc}$) using the SHAM procedure, while that of PATCHY resides at the boundary of 2σ due to using the different Ω_m . For scales $s < 4 h^{-1} \text{Mpc}$, COLA and GADGET deviate slightly more than 2σ , since this scale exceeds the SHAM fitting range.

are relevant for the χ^2 calculations [refer to Eq. (7)]. The results are shown in Fig. 12 to illustrate the normalized MCFs, $s^2 \hat{W}_0(s)$ (top) and $\hat{W}_0/\hat{W}_0^{\text{CMASS}}$ (bottom). The error bars represent the 2σ errors, estimated using the PATCHY mocks.

It is evident that the normalized statistics for the COLA and GADGET mocks exhibit good agreement under the same cosmological parameter $\Omega_m = 0.2951$. Both COLA and GADGET simulations closely match the CMASS data within approximately 2σ level. These statistics from PATCHY are also in agreement with those from the CMASS data at the 2σ level. Both the COLA and GADGET simulations show one point beyond 2σ in all four cases, essentially due to the cosmic variance. Thus, we observe a good consistency between COLA and the high-accuracy GADGET simulations, indicating that COLA adequately ensures the reliability of our analysis.

For the stage-III surveys such as SDSS, the accuracy of COLA is sufficient given the associated statistical error of the data sample. However, for the ongoing stage-IV surveys, the statistical error of the sample is greatly

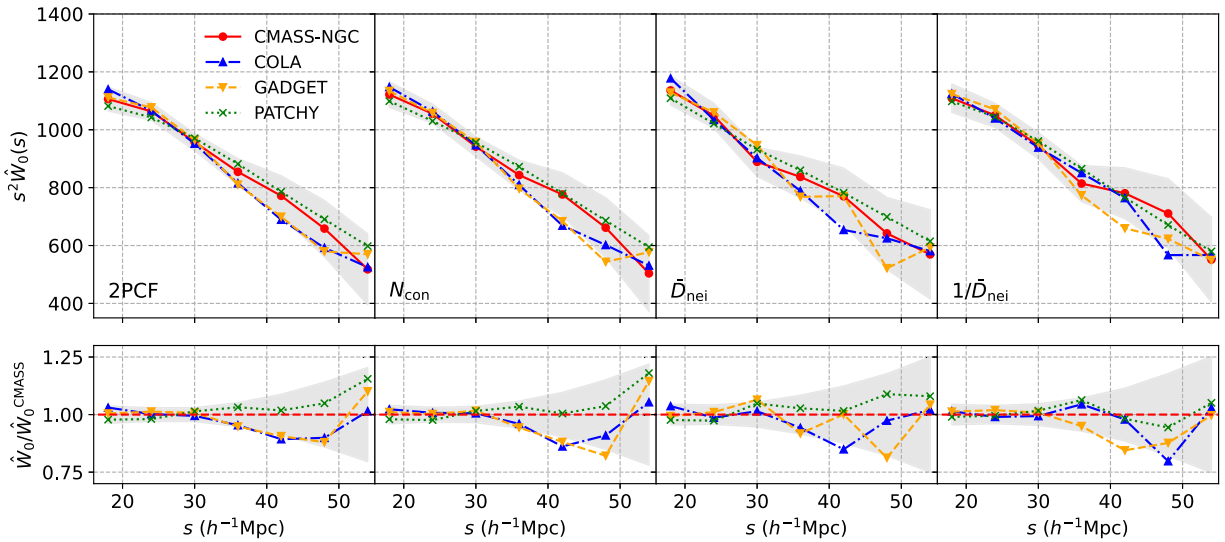


FIG. 12. Similar to Fig. 11, but for the normalized MCFs, $s^2 \hat{W}_0(s)$ (top) and $\hat{W}_0/\hat{W}_0^{\text{CMASS}}$ (bottom) in the range of $s \in [15, 57] h^{-1} \text{Mpc}$, which are relevant for χ^2 calculations [Eq. (7)]. The error bars represent 2σ errors estimated by the PATCHY mocks. COLA and GADGET mocks exhibit good agreement under $\Omega_m = 0.2951$, matching CMASS data within about 2σ level. For all four cases, there is one point slightly exceeding 2σ in both COLA and GADGET, which can be explained by the cosmic variance.

reduced. In this case, it is necessary to improve the work of [50] to ensure that the COLA achieves a higher degree of accuracy. Alternatively, one can directly utilize large-scale,

high-accuracy cosmological N-body simulations that match the stage-IV surveys such as AbacusSummit [75] and Uchuu [76].

- [1] J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay, *Astrophys. J.* **304**, 15 (1986).
- [2] V. de Lapparent, M. J. Geller, and J. P. Huchra, *Astrophys. J.* **302**, L1 (1985).
- [3] J. P. Huchra, L. M. Macri, K. L. Masters, T. H. Jarrett, P. Berlind, M. Calkins, A. C. Crook, R. Cutri, P. Erdoğdu, E. Falco *et al.*, *Astrophys. J. Suppl. Ser.* **199**, 26 (2012).
- [4] M. Tegmark *et al.*, *Astrophys. J.* **606**, 702 (2004).
- [5] L. Guzzo, M. Scodreggio, B. Garilli, B. R. Granett, A. Fritz, U. Abbas, C. Adami, S. Arnouts, J. Bel, M. Bolzonella *et al.*, *Astron. Astrophys.* **566**, A108 (2014).
- [6] J. F. Suárez-Pérez, Y. Camargo, X.-D. Li, and J. E. Forero-Romero, *Astrophys. J.* **922**, 204 (2021).
- [7] D. G. York, J. Adelman, J. E. Anderson, Jr., S. F. Anderson, J. Annis, N. A. Bahcall, J. A. Bakken, R. Barkhouser, S. Bastian, E. Berman *et al.*, *Astron. J.* **120**, 1579 (2000).
- [8] D. J. Eisenstein, I. Zehavi, D. W. Hogg, R. Scoccimarro, M. R. Blanton, R. C. Nichol, R. Scranton, H. Seo, M. Tegmark, Z. Zheng *et al.*, *Astrophys. J.* **633**, 560 (2005).
- [9] W. J. Percival, S. Cole, D. J. Eisenstein, R. C. Nichol, J. A. Peacock, A. C. Pope, and A. S. Szalay, *Mon. Not. R. Astron. Soc.* **381**, 1053 (2007).
- [10] L. Anderson, E. Aubourg, S. Bailey, F. Beutler, V. Bhardwaj, M. Blanton, A. S. Bolton, J. Brinkmann, J. R. Brownstein, A. Burden *et al.*, *Mon. Not. R. Astron. Soc.* **427**, 3435 (2012).
- [11] A. G. Sánchez, C. G. Scóccola, A. J. Ross, W. Percival, M. Manera, F. Montesano, X. Mazzalay, A. J. Cuesta, D. J. Eisenstein, E. Kazin *et al.*, *Mon. Not. R. Astron. Soc.* **425**, 415 (2012).
- [12] A. G. Sánchez, E. A. Kazin, F. Beutler, C.-H. Chuang, A. J. Cuesta, D. J. Eisenstein, M. Manera, F. Montesano, R. C. Nichol, N. Padmanabhan *et al.*, *Mon. Not. R. Astron. Soc.* **433**, 1202 (2013).
- [13] L. Anderson, E. Aubourg, S. Bailey, F. Beutler, V. Bhardwaj, M. Blanton, A. S. Bolton, J. Brinkmann, J. R. Brownstein, A. Burden *et al.*, *Mon. Not. R. Astron. Soc.* **441**, 24 (2014).
- [14] A. J. Ross, L. Samushia, C. Howlett, W. J. Percival, A. Burden, and M. Manera, *Mon. Not. R. Astron. Soc.* **449**, 835 (2015).
- [15] F. Beutler, H.-J. Seo, A. J. Ross, P. McDonald, S. Saito, A. S. Bolton, J. R. Brownstein, C.-H. Chuang, A. J. Cuesta, D. J. Eisenstein *et al.*, *Mon. Not. R. Astron. Soc.* **464**, 3409 (2017).
- [16] A. G. Sánchez, J. N. Grieb, S. Salazar-Albornoz, S. Alam, F. Beutler, A. J. Ross, J. R. Brownstein, C.-H. Chuang, A. J. Cuesta, D. J. Eisenstein *et al.*, *Mon. Not. R. Astron. Soc.* **464**, 1493 (2017).
- [17] S. Alam, M. Ata, S. Bailey, F. Beutler, D. Bizyaev, J. A. Blazek, A. S. Bolton, J. R. Brownstein, A. Burden, C.-H. Chuang *et al.*, *Mon. Not. R. Astron. Soc.* **470**, 2617 (2017).

- [18] C.-H. Chuang, F.-S. Kitaura, Y. Liang, A. Font-Ribera, C. Zhao, P. McDonald, and C. Tao, *Phys. Rev. D* **95**, 063528 (2017).
- [19] M. Colless, G. Dalton, S. Maddox, W. Sutherland, P. Norberg, S. Cole, J. Bland-Hawthorn, T. Bridges, R. Cannon, C. Collins *et al.*, *Mon. Not. R. Astron. Soc.* **328**, 1039 (2001).
- [20] F. Beutler, C. Blake, M. Colless, D. H. Jones, L. Staveley-Smith, G. B. Poole, L. Campbell, Q. Parker, W. Saunders, and F. Watson, *Mon. Not. R. Astron. Soc.* **423**, 3430 (2012).
- [21] F. Beutler, C. Blake, M. Colless, D. H. Jones, L. Staveley-Smith, L. Campbell, Q. Parker, W. Saunders, and F. Watson, *Mon. Not. R. Astron. Soc.* **416**, 3017 (2011).
- [22] D. Parkinson, S. Riemer-Sørensen, C. Blake, G. B. Poole, T. M. Davis, S. Brough, M. Colless, C. Contreras, W. Couch, S. Croom *et al.*, *Phys. Rev. D* **86**, 103518 (2012).
- [23] A. Aghamousa, J. Aguilar, S. Ahlen, S. Alam, L. E. Allen, C. Allende Prieto, J. Annis, S. Bailey, C. Balland *et al.* (DESI Collaboration), [arXiv:1611.00036](https://arxiv.org/abs/1611.00036).
- [24] B. Abareshi, J. Aguilar, S. Ahlen, S. Alam, D. M. Alexander, R. Alfarsy, L. Allen, C. Allende Prieto, O. Alves *et al.* (DESI Collaboration), *Astrophys. J.* **164**, 207 (2022).
- [25] Y. Gong, X. Liu, Y. Cao, X. Chen, Z. Fan, R. Li, X.-D. Li, Z. Li, X. Zhang, and H. Zhan, *Astrophys. J.* **883**, 203 (2019).
- [26] F. Fang, J. Forero-Romero, G. Rossi, X.-D. Li, and L.-L. Feng, *Mon. Not. R. Astron. Soc.* **485**, 5276 (2019).
- [27] M. Davis, G. Efstathiou, C. S. Frenk, and S. D. M. White, *Astrophys. J.* **292**, 371 (1985).
- [28] A. Klypin and J. Holtzman, [arXiv:astro-ph/9712217](https://arxiv.org/abs/astro-ph/9712217).
- [29] V. Springel, S. D. M. White, G. Tormen, and G. Kauffmann, *Mon. Not. R. Astron. Soc.* **328**, 726 (2001).
- [30] S. R. Knollmann and A. Knebe, *Astrophys. J. Suppl. Ser.* **182**, 608 (2009).
- [31] O. Hahn, C. Porciani, C. M. Carollo, and A. Dekel, *Mon. Not. R. Astron. Soc.* **375**, 489 (2007).
- [32] J. E. Forero-Romero, Y. Hoffman, S. Gottlöber, A. Klypin, and G. Yepes, *Mon. Not. R. Astron. Soc.* **396**, 1815 (2009).
- [33] Y. Hoffman, O. Metuki, G. Yepes, S. Gottlöber, J. E. Forero-Romero, N. I. Libeskind, and A. Knebe, *Mon. Not. R. Astron. Soc.* **425**, 2049 (2012).
- [34] J. E. Forero-Romero, S. Contreras, and N. Padilla, *Mon. Not. R. Astron. Soc.* **443**, 1090 (2014).
- [35] B. Reid, S. Ho, N. Padmanabhan, W. J. Percival, J. Tinker, R. Tojeiro, M. White, D. J. Eisenstein, C. Maraston, A. J. Ross *et al.*, *Mon. Not. R. Astron. Soc.* **455**, 1553 (2015).
- [36] A. Klypin, G. Yepes, S. Gottlöber, F. Prada, and S. Hess, *Mon. Not. R. Astron. Soc.* **457**, 4340 (2016).
- [37] M. V. García-Alvarado, J. E. Forero-Romero, and X.-D. Li, *Mon. Not. R. Astron. Soc.* **498**, L145 (2020).
- [38] S. Genel, M. Vogelsberger, V. Springel, D. Sijacki, D. Nelson, G. Snyder, V. Rodriguez-Gomez, P. Torrey, and L. Hernquist, *Mon. Not. R. Astron. Soc.* **445**, 175 (2014).
- [39] M. Vogelsberger, S. Genel, V. Springel, P. Torrey, D. Sijacki, D. Xu, G. Snyder, S. Bird, D. Nelson, and L. Hernquist, *Nature (London)* **509**, 177 (2014).
- [40] D. Nelson, V. Springel, A. Pillepich, V. Rodriguez-Gomez, P. Torrey, S. Genel, M. Vogelsberger, R. Pakmor, F. Marinacci, R. Weinberger *et al.*, *Comput. Astrophys. Cosmol.* **6**, 2 (2019).
- [41] C. G. Sabiu, D. F. Mota, C. Llinares, and C. Park, *Astron. Astrophys.* **592**, A38 (2016).
- [42] Z. Slepian, D. J. Eisenstein, F. Beutler, C.-H. Chuang, A. J. Cuesta, J. Ge, H. Gil-Marín, S. Ho, F.-S. Kitaura, C. K. McBride *et al.*, *Mon. Not. R. Astron. Soc.* **468**, 1070 (2017).
- [43] C. G. Sabiu, B. Hoyle, J. Kim, and X.-D. Li, *Astrophys. J. Suppl. Ser.* **242**, 29 (2019).
- [44] M. White, *J. Cosmol. Astropart. Phys.* **11** (2016) 057.
- [45] Y. Yang, H. Miao, Q. Ma, M. Liu, C. G. Sabiu, J. Forero-Romero, Y. Huang, L. Lai, Q. Qian, Y. Zheng *et al.*, *Astrophys. J.* **900**, 6 (2020).
- [46] L. M. Lai, J. C. Ding, X. L. Luo, Y. Z. Yang, Z. H. Wang, K. S. Liu, G. F. Liu, X. Wang, Y. Zheng, Z. Y. Li *et al.*, [arXiv:2312.03244](https://arxiv.org/abs/2312.03244).
- [47] S. Satpathy, R. Croft, S. Ho, and B. Li, *Mon. Not. R. Astron. Soc.* **484**, 2148 (2019).
- [48] X. Xiao, Y. Yang, X. Luo, J. Ding, Z. Huang, X. Wang, Y. Zheng, C. G. Sabiu, J. Forero-Romero, H. Miao *et al.*, *Mon. Not. R. Astron. Soc.* **513**, 595 (2022).
- [49] S. Tassev, M. Zaldarriaga, and D. J. Eisenstein, *J. Cosmol. Astropart. Phys.* **06** (2013) 036.
- [50] J. Ding, S. Li, Y. Zheng, X. Luo, L. Zhang, and X.-D. Li, *Astrophys. J. Suppl. Ser.* **270**, 25 (2024).
- [51] F.-S. Kitaura, S. Rodríguez-Torres, C.-H. Chuang, C. Zhao, F. Prada, H. Gil-Marín, H. Guo, G. Yepes, A. Klypin, C. G. Scóccola *et al.*, *Mon. Not. R. Astron. Soc.* **456**, 4156 (2016).
- [52] D. J. Eisenstein, D. H. Weinberg, E. Agol, H. Aihara, C. Allende Prieto, S. F. Anderson, J. A. Arns, É. Aubourg, S. Bailey, E. Balbinot *et al.*, *Astron. J.* **142**, 72 (2011).
- [53] A. S. Bolton, D. J. Schlegel, É. Aubourg, S. Bailey, V. Bhardwaj, J. R. Brownstein, S. Burles, Y.-M. Chen, K. Dawson, D. J. Eisenstein *et al.*, *Astrophys. J.* **144**, 144 (2012).
- [54] K. S. Dawson, D. J. Schlegel, C. P. Ahn, S. F. Anderson, É. Aubourg, S. Bailey, R. H. Barkhouser, J. E. Bautista, A. Beifiori, A. A. Berlind *et al.*, *Astron. J.* **145**, 10 (2012).
- [55] D. J. Eisenstein, J. Annis, J. E. Gunn, A. S. Szalay, A. J. Connolly, R. C. Nichol, N. A. Bahcall, M. Bernardi, S. Burles, F. J. Castander *et al.*, *Astron. J.* **122**, 2267 (2001).
- [56] P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. J. Banday, R. B. Barreiro, J. G. Bartlett, N. Bartolo *et al.*, *Astron. Astrophys.* **594**, A13 (2016).
- [57] F.-S. Kitaura, S. Rodríguez-Torres, C.-H. Chuang, C. Zhao, F. Prada, H. Gil-Marín, H. Guo, G. Yepes, A. Klypin, C. G. Scóccola *et al.*, *Mon. Not. R. Astron. Soc.* **456**, 4156 (2016).
- [58] F. S. Kitaura, G. Yepes, and F. Prada, *Mon. Not. R. Astron. Soc.* **439**, L21 (2013).
- [59] S. A. Rodríguez-Torres, C.-H. Chuang, F. Prada, H. Guo, A. Klypin, P. Behroozi, C. H. Hahn, J. Comparat, G. Yepes, A. D. Montero-Dorta *et al.*, *Mon. Not. R. Astron. Soc.* **460**, 1173 (2016).
- [60] H. Edelsbrunner, D. G. Kirkpatrick, and R. Seidel, *IEEE Trans. Inf. Theory* **29**, 551 (1983).
- [61] N. Amenta, M. Bern, and D. Eppstein, *Graph. Models Image Process.* **60**, 125 (1998).
- [62] W. Zhang and I. King, *Proceedings of the 9th International Conference on Neural Information Processing, ICONIP '02* (IEEE, Singapore, 2002), pp. 1423–1427, [10.1109/ICONIP.2002.1202855](https://doi.org/10.1109/ICONIP.2002.1202855).

- [63] J. D. Barrow, S. P. Bhavsar, and D. H. Sonoda, *Mon. Not. R. Astron. Soc.* **216**, 17 (1985).
- [64] T. Sousbie, *Mon. Not. R. Astron. Soc.* **414**, 350 (2011).
- [65] P. Pranav, R. van de Weygaert, G. Vegter, B. J. T. Jones, R. J. Adler, J. Feldbrugge, C. Park, T. Buchert, and M. Kerber, *Mon. Not. R. Astron. Soc.* **485**, 4167 (2019).
- [66] J. Feldbrugge, R. van de Weygaert, J. Hidding, and J. Feldbrugge, *J. Cosmol. Astropart. Phys.* **05** (2018) 027.
- [67] D. G. Kirkpatrick and J. D. Radke, in *Computational Geometry*, edited by G. T. Toussaint, Machine Intelligence and Pattern Recognition Vol. 2 (North-Holland, Amsterdam, 1985), pp. 217–248, 10.1016/B978-0-444-87806-9.50013-X.
- [68] C. D. Correa and P. Lindstrom, *Int. J. Uncertainty Quantif.* **3**, 187 (2013).
- [69] S. Landy and A. Szalay, *Astrophys. J.* **412**, 64 (1993).
- [70] X.-D. Li, C. Park, C. G. Sabiu, H. Park, D. H. Weinberg, D. P. Schneider, J. Kim, and S. E. Hong, *Astrophys. J.* **832**, 103 (2016).
- [71] V. Springel, N. Yoshida, and S. D. White, *New Astron.* **6**, 79 (2001).
- [72] V. Springel, *Mon. Not. R. Astron. Soc.* **364**, 1105 (2005).
- [73] V. Springel, R. Pakmor, O. Zier, and M. Reinecke, *Mon. Not. R. Astron. Soc.* **506**, 2871 (2021).
- [74] G. Xu, *Astrophys. J. Suppl. Ser.* **98**, 355 (1995).
- [75] N. A. Maksimova, L. H. Garrison, D. J. Eisenstein, B. Hadzhiyska, S. Bose, and T. P. Satterthwaite, *Mon. Not. R. Astron. Soc.* **508**, 4017 (2021).
- [76] T. Ishiyama, F. Prada, A. A. Klypin, M. Sinha, R. B. Metcalf, E. Jullo, B. Altieri, S. A. Cora, D. Croton, S. de la Torre *et al.*, *Mon. Not. R. Astron. Soc.* **506**, 4210 (2021).