



CHORUS

This is the accepted manuscript made available via CHORUS. The article has been published as:

Percolation analysis for cosmic web with discrete points

Jiajun Zhang, Dalong Cheng, and Ming-Chung Chu

Phys. Rev. D **97**, 023534 — Published 29 January 2018

DOI: [10.1103/PhysRevD.97.023534](https://doi.org/10.1103/PhysRevD.97.023534)

New leaves of the tree: percolation analysis for cosmic web with discrete points

Jiajun Zhang

*School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai, China and
Department of Physics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong*

Dalong Cheng

School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai, China

Ming-Chung CHU

Department of Physics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong

Percolation analysis has long been used to quantify the connectivity of the cosmic web. Most of the previous work is based on density fields on grids. By smoothing into fields, we lose information about galaxy properties like shape or luminosity. The lack of mathematical modelling also limits our understanding for the percolation analysis. In order to overcome these difficulties, we have studied percolation analysis based on discrete points. Using a Friends-of-Friends (FoF) algorithm, we generate the $S - bb$ relation, between the fractional mass of the largest connected group (S) and the FoF linking length (bb). We propose a new model, the Probability Cloud Cluster Expansion Theory (PCCET) to relate the $S - bb$ relation with correlation functions. We show that the $S - bb$ relation reflects a combination of all orders of correlation functions. Using N-body simulation, we find that the $S - bb$ relation is robust against redshift distortion and incompleteness in observation. From the Bolshoi simulation, with Halo Abundance Matching (HAM), we have generated a mock galaxy catalogue. Good matching of the projected two-point correlation function with observation is confirmed. However, comparing the mock catalogue with the latest galaxy catalogue from SDSS DR12, we have found significant differences in their $S - bb$ relations. This indicates that the mock galaxy catalogue cannot accurately retain higher order correlation functions than the two-point correlation function, which reveals the limit of the HAM method. As a new measurement, $S - bb$ relation is applicable to a wide range of data types, fast to compute, robust against redshift distortion and incompleteness, and it contains information of all orders of correlation functions.

Keywords: methods: numerical – galaxies: statistics – cosmology: theory – cosmology: large-scale structure of Universe – cosmology: cosmic web.

I. INTRODUCTION

Cellular or clumpy? That is the question. In the last century, when the first generation of redshift surveys were available, people were confused by the filamentary structure of galaxy distribution [1, 2]. The existence of the cosmic web was qualitatively confirmed, but a quantitative description of it was not available until Zeldovich’s percolation analysis [3–5]. We classify the percolation analysis for the cosmic web into two branches. The first one, dealing with discrete points, is called continuum percolation, whereas the second one, dealing with the density field on grids, is called site percolation. Continuum percolation analysis was also called cluster analysis in 1980s, because the same technique was used to define galaxy clusters. Cluster analysis can clearly tell the difference between a clumpy structure and web-like pattern [6]. However, Dekel & West [7] concluded that cluster analysis was insensitive to cosmological parameters, and it was also non-trivially affected by volume and point density. Thereafter, the site percolation analysis became popular, and properties of site percolation in large scale density fields have been studied in detail [8–16]. By smoothing points into field, the technique has also been used to look for properties of galaxy distribution [17–19]. We would like to point out that direct comparison between the percolation properties of density field and galaxy distribution is not fair, because galaxies don’t trace the density field perfectly. If we can generate a mock catalogue of galaxies and compare it with a real redshift survey, the comparison will be fair and the result will give us some useful insights in properties of the cosmic web.

On the other hand, the establishment of the Λ CDM model and rapid development of N-body simulation techniques provide us a general picture of the formation of the cosmic web [20–22]. We are now confident about the success of the Λ CDM model in large scales. However the three-point, four-point or even N-point correlation functions should also be tested [23–27], though they are all very difficult to calculate and model in theory. In principle, the two-point correlation function costs $O(N^2)$ calculations, and the three-point correlation function costs $O(N^3)$ calculations, so on and so forth. Although we have better ways to optimize the calculation, higher order correlation functions still require unbearable calculation resources. Thus, several methods to characterize the cosmic web have been proposed, such as Minkowski Functional, Minimum Spanning Tree, and Genus Analysis [28–31]. These methods focus on the morphology and topology of the cosmic web.

The comparison between simulation and observation is not straight forward. We are still lacking of knowledge about galaxy formation and evolution, and matching galaxies with simulated dark halos heavily depends on model [32–35]. Surprisingly, Halo Abundance Matching (HAM), a newly developed technique based on simple arguments, generates mock galaxy catalogues successfully [36–38]. HAM follows the halo mass, halo circular velocity, halo potential well, galaxy circular velocity, galaxy luminosity, galaxy stellar mass, independent of galaxy formation model, and HAM is easy to understand.

Shandarin et al. [39] developed and studied the percolation analysis, treating voids and superclusters on an equal footing and finding the nature and nurture of the formation of cosmic web. Using a similar method, the SDSS galaxy sample has been tested with the percolation analysis [19]. However we are still far from understanding the percolation transition in the cosmic web. Density estimation is a prior step for site percolation analysis, but estimating the density field from galaxy observation is not easy. We should take into account galaxies’ different properties in the percolation analysis.

Based on these considerations, we used the continuum percolation analysis to calculate the $S - bb$ relation, where bb stands for the FoF linking length and S stands for the mass fraction of the largest group for the given linking length bb [40]. It is quite similar to the largest cluster statistics in previous percolation analyses [7]. However, we find that the continuum percolation analysis can distinguish among cosmological models with high resolution simulations, whose resolution was not available in 1985. We propose the Probability Could Cluster Expansion Theory (PCCET), to calculate the $S - bb$ relation theoretically. We measure the $S - bb$ relation in our mock galaxy catalogue and compare it with a real galaxy catalogue. We find that HAM cannot reproduce the $S - bb$ relation as well as the projected two-point correlation function.

II. METHODOLOGY

A. S-bb Relation and Percolation Transition

For any discrete point distribution, we can always apply the FoF algorithm to find a group of points. The points can be massive points in N-body simulations, halos in simulated halo catalogues or galaxy catalogues in redshift space. We define S_1 and S_2 as:

$$S_1(bb) = N_{largest}/N_{total}, \quad (1)$$

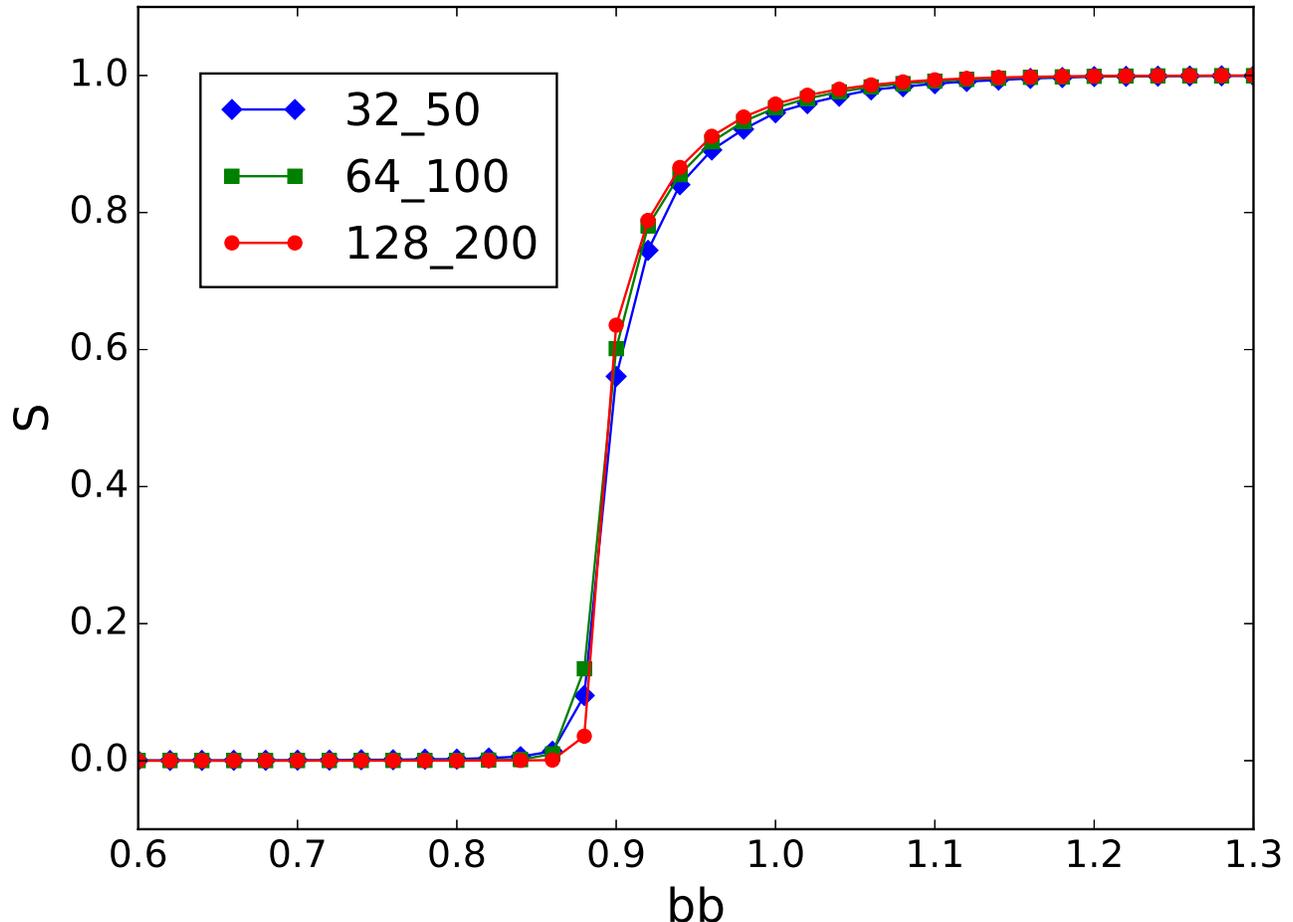


FIG. 1. Three sets of $S - bb$ relation for random distributions are shown here, corresponding to 32^3 , 64^3 and 128^3 particles in boxes with $50h^{-1}\text{Mpc}$, $100h^{-1}\text{Mpc}$, and $200h^{-1}\text{Mpc}$ sides. Note that these three curves are degenerate and the transition threshold is around $bb = 0.88$.

$$S_2(bb) = N_{\text{second-largest}}/N_{\text{total}}, \quad (2)$$

where N_{total} is the total number of points, N_{largest} ($N_{\text{second-largest}}$) is the number of points in the largest (second largest) group found by FoF, and normally, we refer to the $S - bb$ relation the function of $S_1(bb)$. It is obvious that $S_1(bb)$ reveals quite similar information as the cluster analysis and site percolation analysis. Our simulation tests are all done using `Gadget2` [41]. Fig. 1 is the $S - bb$ relation for a random distribution as a comparison for later figures. It is not surprising that all three sets of data corresponding to different box sizes and number of particles give the same $S_1(bb)$ and the transition threshold is around $bb = 0.88$.

We define the transition threshold as the bb where $S_2(bb)$ gets the maximum value. Another definition of the transition threshold was also proposed in [8]; in our case, it can be defined as the bb where the mass-weighted sum of the number of connected structures, $\mu^2 = \frac{\sum_n N_n}{N_{\text{groups}}}$, where N_n is the number of particles in the n^{th} group and N_{groups} is the total number of connected structures, gets the maximum value.

We measured the S_1 , S_2 and μ^2 using the halo catalogue of the MDR1 simulation[42] with haloes with mass larger than $10^{12}M_{\odot}$. The MDR1 simulation was performed in a box with $1h^{-1}\text{Gpc}$ sides, and so we separated the sample into 64 subsamples of boxes with $250h^{-1}\text{Mpc}$ sides to estimate the cosmic variance, shown as the errorbars in Fig. 2. The transition threshold bb_c is 0.74 and the power index α given by $S_1 = (bb - bb_c)^\alpha$ was measured to be $\alpha = 0.46$, which is consistent with [8].

As shown in Fig. 2, both definitions of the transition threshold are good, and their FWHM are similar. We will discuss the meaning of the transition threshold in the definition of S_2 later for easier understanding. When the

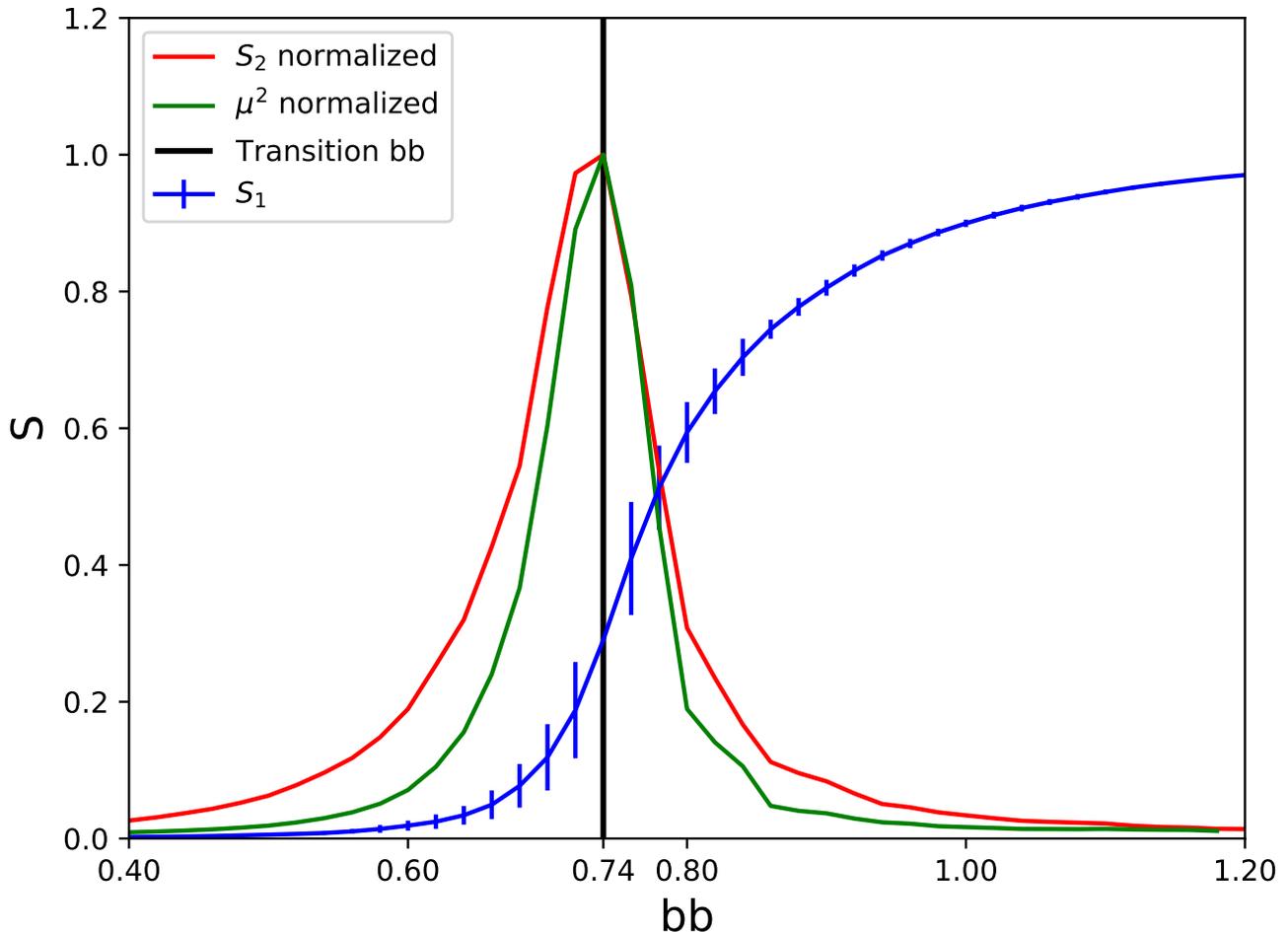


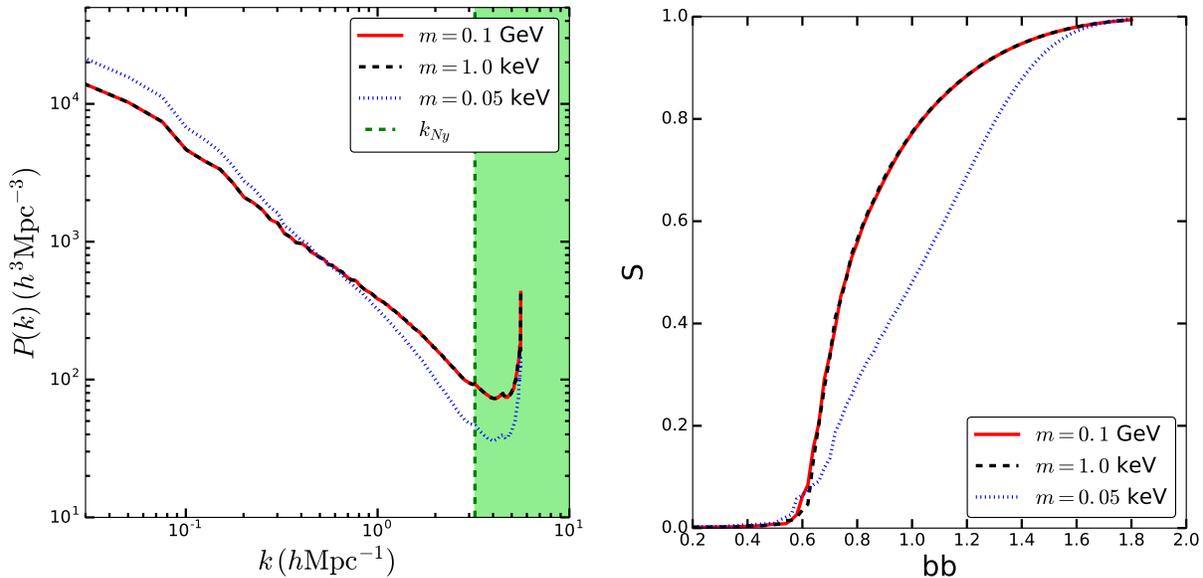
FIG. 2. The $S - bb$ relation measured from the MDR1 simulation [42]. The blue line is S_1 , the red line is the normalized S_2 , and the green line is the normalized μ^2 . The transition threshold is shown as the vertical black line, at $bb = 0.74$ in this plot. The sharp peak of S_2 reveals the nature of percolation, that most massive structures combine into the largest structure which show up beyond the size of superclusters. The FWHM of μ^2 and S_2 are similar, but the FWHM of μ^2 is $\sim 10\%$ smaller than that of S_2 . The shape of S_1 is very much different from the random sample shown in Fig. 1. It shows clearly that the cosmic web is not random.

percolation transition happens, the largest structure grows tremendously as bb gets larger, and the second largest structure reaches its most massive point. After that, the mass of the second largest structure becomes smaller, not because the structure shrinks, but because the original second largest structure is connected to the largest one while another structure is recognized as the new second largest one. This connection and replacement reveals the nature of percolation: the largest structure becomes the one and the only one dominant structure, connecting most of the high density regions together and leaving only the far apart particles in voids alone. Thus, the definition of transition threshold based on either S_2 or μ^2 is appropriate.

While the Λ CDM (cold dark matter) model is not clearly distinguishable from the Λ WDM (warm dark matter) model, the Λ HDM (hot dark matter) model is clearly ruled out by observation of large scale structures. We would like to test our percolation analysis with these three models. The $S - bb$ relations are expected to be indistinguishable for Λ CDM and Λ WDM models, but those for Λ HDM and Λ CDM models should be clearly different. This test is the bench mark test of percolation analysis. We generate the initial conditions with 2LPTic [43] and run the simulation from redshift 49 to redshift zero using the N-body simulation code Gadget2[41]. The transfer function and additional thermal velocity for HDM and WDM are given by [44]. These three simulations share the same realization. Some basic parameters for the simulations are given in Table. I. As shown in Fig. 3, the HDM model is quite different from CDM and WDM models in their matter power spectra and $S - bb$ relations. The matter power spectrum is calculated by the publicly available code ComputePk[46]. It is well accepted that the HDM model has been ruled

TABLE I. Basic parameters for three simulations. EH is the power spectrum given by [45].

Model	$m(\text{keV})$	IC	N	$L(h^{-1}\text{Mpc})$
CDM	10^5	EH	256^3	250
WDM	1	EH	256^3	250
HDM	0.05	EH	256^3	250

FIG. 3. Comparison of the matter power spectra and $S - bb$ relations for Hot (red curves), Warm (green curves) and Cold (blue curves) Dark Matter models. The green shaded region on the left panel indicates the resolution limit of the simulations.

out by measuring the power spectrum in both CMB [47] and large scale structures [48]. We have shown here that the difference between HDM model and CDM model is significant enough in the $S - bb$ relation to be used as a supplementary. On the other hand, the unique advantage of using the $S - bb$ relation will be discussed in the following sections.

B. Probability Cloud Cluster Expansion Theory

The percolation analysis has been used for a long time to study the cosmic web. For site percolation there are analogies to similar questions in many other fields [49–52]. But for discrete points, there is still no available theory to link percolation with basic cosmological parameters. Here we propose the Probability Cloud Cluster Expansion Theory (PCCET) to connect the $S - bb$ relation with correlation functions. The idea is as follows:

1. Consider a box with only one particle and take it as the only member in the largest group. So the $S - bb$ relation will be a constant line $S = 1$.
2. Add in another particle randomly, the probability of linking these two particles in a single group $P(bb)$ being monotonically related to the linking length bb . The $S - bb$ relation will be a step function jumping from $S_1 = 0.5$ to $S_1 = 1.0$ at a certain transition value of P : $S = \frac{1}{2} + \frac{1}{2}P$.
3. Add in the third particle randomly, the probability of linking any two of them being still P and the probability of linking all three of them shall be P^2 . While from a certain particle as the starting point of the group, the probability of linking the other two particles shall be $2P$ but deducting P^2 to avoid over counting. The $S - bb$ relation will then be a three-step function, with steps at $1/3$, $2/3$, and 1 : $S = \frac{1}{3} + \frac{1}{3}P + \frac{1}{3}P(2P - P^2)$.
4. Continuing on, we consider more and more particles, taking care of over counting and over deducting, and we finally find the relation $S(P)$ when the number of particles goes to infinity. We hide all information about bb in the function $P(bb)$.

$$P = \begin{cases} \frac{4}{3}\pi\left(\frac{bb}{2}\right)^3, & 0 \leq bb \leq 1, \\ \frac{4}{3}\pi\left(\frac{bb}{2}\right)^3 - 6\pi\left(\frac{bb-1}{2}\right)^2\left(\frac{bb}{2} - \frac{bb-1}{6}\right), & 1 < bb \leq \sqrt{2}, \\ 1 - \int_0^{1-\sqrt{bb^2-2}} dx \int_0^{\sqrt{bb^2-1-(x-1)^2}} dy \int_0^{1-\sqrt{bb^2-(x-1)^2-(y-1)^2}} dz, & \sqrt{2} < bb \leq \sqrt{3}, \\ 1, & bb > \sqrt{3}. \end{cases} \quad (4)$$

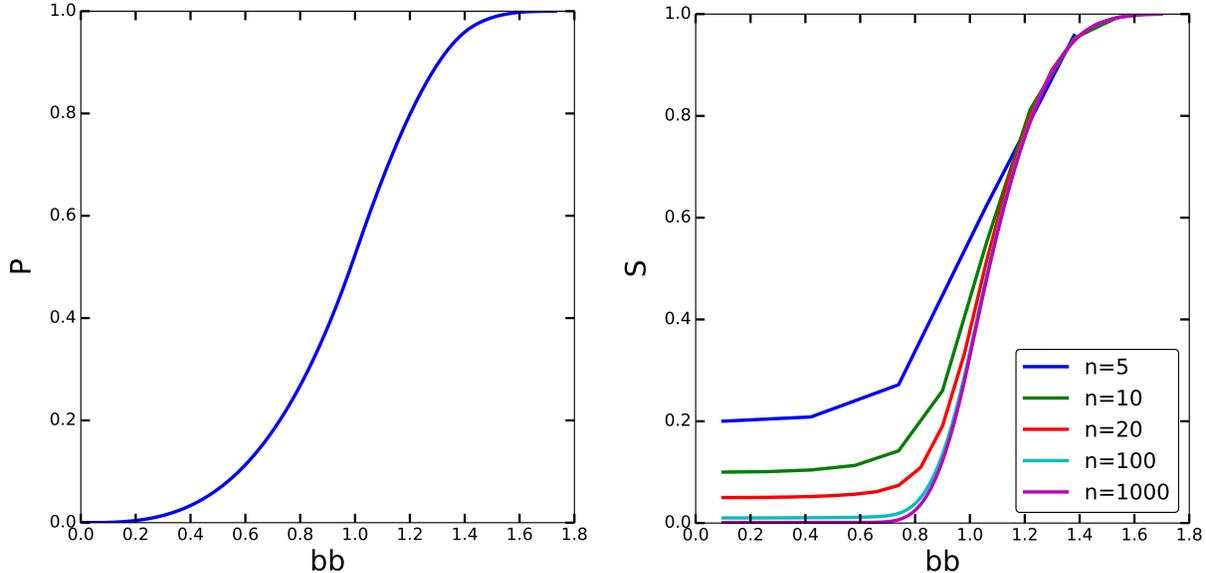


FIG. 4. The $P - bb$ relation given in Eq. 4 is shown on the left. The $S - bb$ relation according to Eq.s (4) and (3) is shown on the right. Notice the differences of the curves with different orders of expansion n given by Eq. (3).

So in general, we can write $S(P)$ as

$$\begin{cases} a_1 = \frac{1}{N}, \\ a_n = a_{n-1}b_{n-1} (n > 1), \\ b_n = 1 - (1 - P)^n, \\ S = \sum_{n=1}^N a_n. \end{cases} \quad (3)$$

This series expansion is the reason why we call it cluster expansion theory. Cluster expansion is often used in condensed matter physics. It can successfully account for phase transitions in some typical problems. We borrow the idea of cluster expansion because of the similarity between percolation transition and phase transition. We have shown that $S(P)$ is monotonically increasing with P and has a limit between 0 and 1. When $N \rightarrow \infty$, $S(P)$ will become the Euler Function. The proof in detail is given in Sec. V A.

In the framework of this theory, a particle cannot be treated as a discrete point. In fact, it is better to take the particle as a representation of the probability cloud. For a random distribution, the probability for a particle to be closer to another particle than distance r is inversely proportional to the volume of a sphere with radius r . However, we need a volume to set as the standard for this probability. Here we take all the particles in the theory as test points. We can first put in the test center. We imagine that every test center has its own cube with periodic boundary conditions. The length of each side of this test cube can be set as the average distance between particles in a real simulation. Then the distance to the center can be simply understood as the linking length bb . If bb is larger than $1/2$, the influence of this test center will reach outside the test box. In such a case, we shall remove the influence region outside the test box and only consider the volume inside the box as the real probability distribution region. Shown in Eq. 4 is the $P - bb$ relation for a random distribution.

From the PCCET, we can find the relation between percolation and correlation function. Because of the cluster expansion, the 2-point correlation function is the correction to P as the probability of linking up 2 points, while the

3-point correlation function is the correction to P^2 as the probability of linking up 2 pairs of points, etc. In general, the n -point correlation function is the correction to $P^{(n-1)}$ as the probability of linking up $n-1$ pairs of points. This shows that percolation actually contains the information of all orders of correlation functions. Our probability cloud cluster expansion theory also shows how every order of correlation function contributes to the $S-bb$ relation, which represents the percolation phenomenon in the cosmic web.

Unfortunately, in the non-linear regime, high order correlation functions are also quite large and not necessarily smaller than the 2-point correlation function. So based on our theory, it is still not possible to calculate the whole $S-bb$ relation analytically. It is important to use N-body simulations to set up constraints for the $S-bb$ relation. After all, PCCET is still the only model explaining percolation. It can be used to find the $S-bb$ relation in the linear-regime and it is also possible that in the future, we may find a way to use it in the non-linear regime to extract high order correlation functions.

C. Halo Abundance Matching

From PCCET, we can see that the $S-bb$ relation contains information of all orders of correlation functions. Thus we would like to see whether it provides new information in comparing simulation with observation. We first need to generate a mock galaxy catalogue and check the projected 2-point correlation function to see whether it matches the real galaxy catalogue. Then we will compare the $S-bb$ relations of the mock and real galaxy catalogues.

Based on the Bolshoi simulation [53, 54] BDM halo merger tree, we follow the work in [36] using Halo Abundance Matching to generate a mock galaxy catalogue. Here is the process:

1. We take the Bolshoi simulation BDM halo catalogue at redshift 0 to extract the final positions and peculiar velocities of the galaxies. We assume that every BDM halo, including the host haloes and subhaloes, can host at most one galaxy at its center.
2. From the merger tree of BDM haloes in the Bolshoi simulation, we find the maximum circular velocity V_{max} in the merger history with the halo circular velocity defined as the maximum value of the circular orbital velocity according to the density distribution of the halo.
3. We adopt the luminosity function from [55], which is of the Schechter form. The luminosity function is:

$$\begin{aligned} \Phi(M)dM &= 0.4 \log_{10} \Phi_* 10^{-0.4(M-M_*)(\alpha+1)} \\ &\times \exp(-10^{-0.4(M-M_*)})dM, \end{aligned} \quad (5)$$

where M is the absolute magnitude and $\Phi_* = 0.0078$, $M_* - 5 \log_{10} h = -20.83$, $\alpha = -1.24$, given by [55] using SDSS6 data.

4. We sort the haloes in the Bolshoi simulation according to their V_{max} . Record every halo's ranking. Thus, we can easily obtain the number density of haloes with V_{max} larger than a certain value by the ranking.
5. The integral of the luminosity function is an incomplete gamma function. We apply an one-one mapping from the number density of haloes with V_{max} larger than a certain value V to the integral of the luminosity function. Thus we obtain an one-one mapping between V and absolute magnitude M .
6. Combining the data together, we get the halo catalogue with the galaxy absolute magnitude. This is also the mock galaxy catalogue with scale $250h^{-1}\text{Mpc}$. In this catalogue, we have positions, peculiar velocities and luminosities for the mock galaxies at redshift 0.

Halo Abundance Matching has been proven quite successful in recreating a galaxy catalogue which can reproduce the observed projected galaxy 2-point correlation function in the bins of absolute magnitude close to the steep transform range in the luminosity function, around $\Phi = \Phi_*$. Many works follow the idea of HAM [37, 38]. There are slightly different methods to obtain a mock catalogue based on the same basic idea of HAM. Comparing to [36], we did not take stochastic scattering into account. Since there are obvious bias to dimmer galaxies in stochastic scattering, we avoid this process. On the other hand, we take maximum circular velocity in the merger history of every halo as its monotonic mapping label. The physical insight of this picture is clear. Gas should fall into the center of a halo much faster than the merger rate, and once the gas has fallen into the halo center, it is so dense that the merger process cannot do much to the central dense region. So the galaxy luminosity depends more on the steepest potential well in the merger history, denoted by the maximum circular velocity V_{max} .

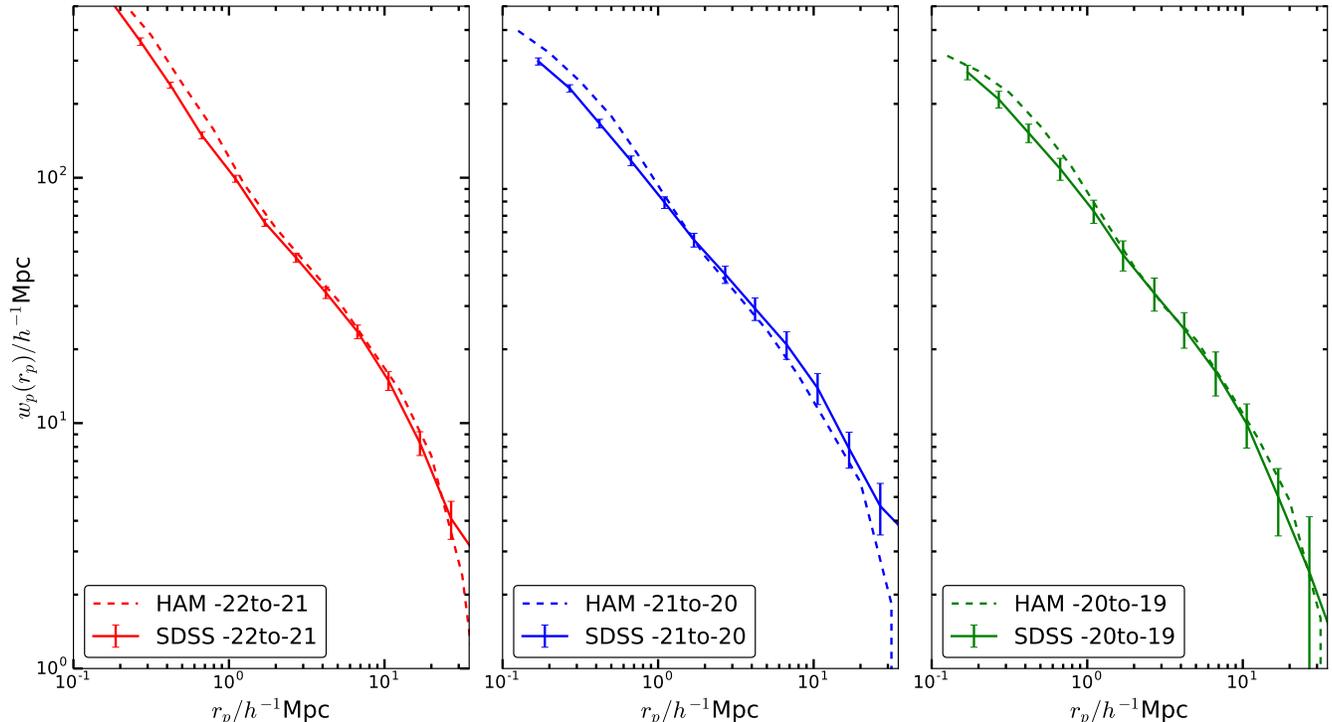


FIG. 5. Comparison of the projected 2-point correlation functions. The solid lines with error bars are extracted from [57], and the dashed lines are given by our measurement on the HAM mock samples. The mock catalogue can reproduce the projected 2-point correlation functions from the observation for all three luminosity ranges.

We follow the paper [56] to calculate the projected 2-point correlation function. Our results are shown in Fig. 5. We have confirmed the previous results given by [36] that the HAM reproduced sample exhibits the same projected 2-point correlation functions as the observation in the luminosity range $[-20,-19]$, $[-21,-20]$ and $[-22,-21]$, with differences mostly within 1 or 2 σ . We take these samples for further percolation analysis comparison. We shall also notice that the bias of the projected 2-point correlation function depends on the luminosity of the galaxy. We shall have a different view of the bias in the $S - bb$ relation.

III. RESULT

A. $S - bb$ Relation Evolution

We are interested in the evolution of the cosmic web, which we study through an Λ CDM cosmological simulation with 512^3 particles in a box with $1h^{-1}\text{Gpc}$ sides, initial conditions generated by BBKS power spectrum [58] and Zel'dovich approximation [59]. We apply $\Omega_m = 0.28$, $\Omega_\Lambda = 0.72$ and $h = 0.7$ for this test. As shown in Fig. 6, the $S - bb$ relation at high redshifts looks similar to that of random distribution (Fig. 1), while as the density contrast increases with time, the $S - bb$ relation at low redshift becomes rather different, which has already been known in many previous works [7, 10].

B. Bias Determination

The error bars of the $S - bb$ relation shown in fig. 7 are determined as follows. Based on Bolshoi simulation [53], we have only one realization of the HAM mock sample. So we use jackknife re-sampling to estimate the error bars. Here are the steps:

1. randomly choose 90% of the HAM mock sample for ten times, generating 10 different mock samples.

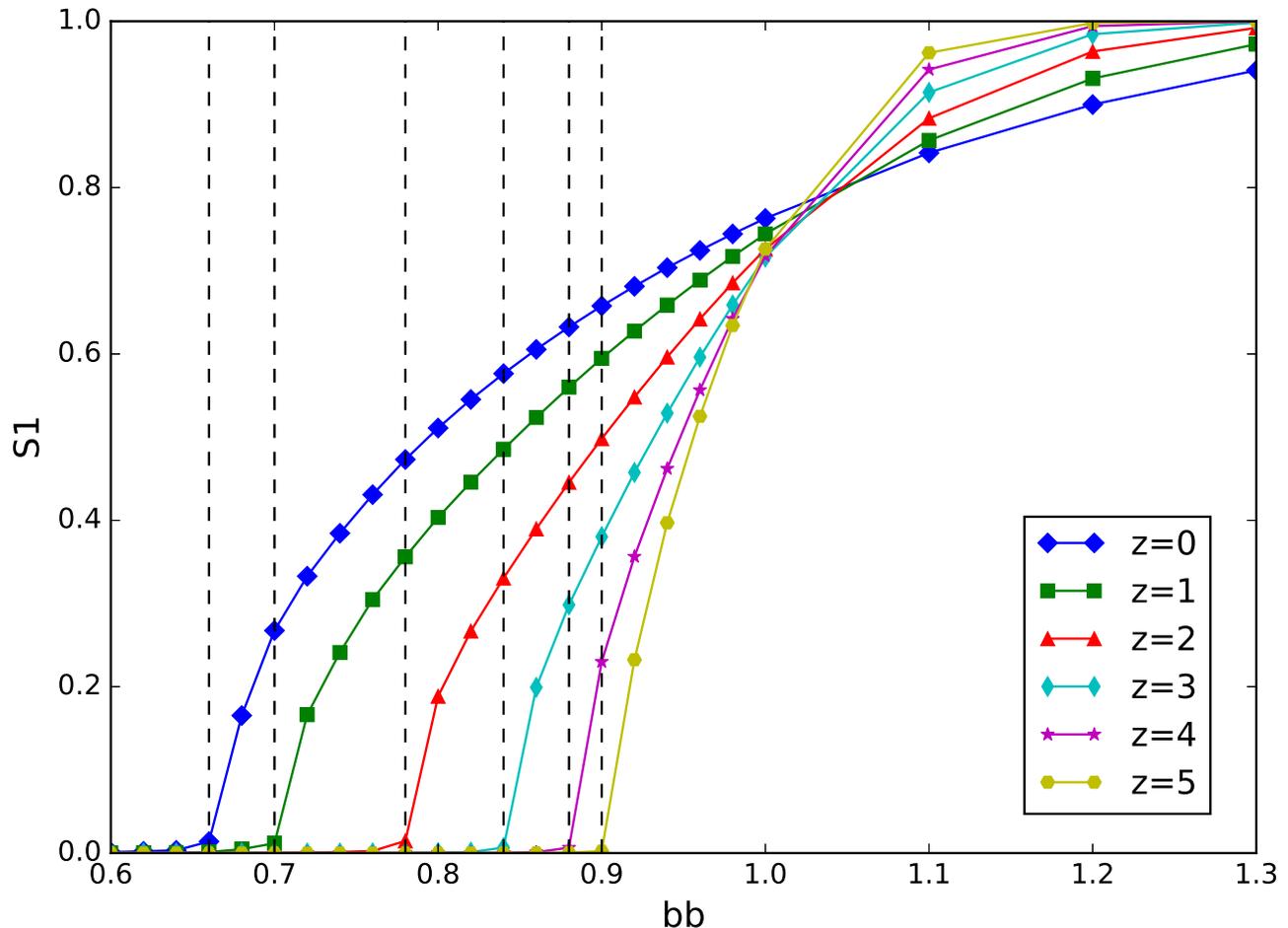


FIG. 6. The redshift (z) dependence of the $S - bb$ relation for Λ CDM cosmological simulation with 512^3 particles in an $1h^{-1}\text{Gpc}$ box. We can see that the transition threshold bb_c gets smaller for lower redshift, which is expected because the 2-point correlation function also becomes larger.

2. Measure the $S - bb$ relations of these 10 mock samples.
3. Take the mean and standard variance of these 10 $S - bb$ relations as the value and error bar.

From Fig. 7 we can see that the lower luminosity samples' transition threshold bb is lower than those of the high luminosity samples. Based on this result, we may guess that the lower luminosity samples have higher bias in the 2-point correlation function than higher luminosity samples. However it is well known that more luminous galaxies have higher bias which seems to be contradictory to our results here. Note that the definition of bb is such that it is normalized to the average distance of points. So we shall look at the 2-point correlation function with the projected distances r_p also normalized to the average distance, which is shown in Fig. 8. The higher the luminosity of the samples are, the lower their normalized 2-point correlation function is. Therefore, a smaller transition threshold is well expected.

Galaxies' real spatial distribution is different from their redshift space distribution due to the peculiar velocities of the galaxies. Using the HAM mock sample data with peculiar velocity, we can study the redshift distortion effect in the $S - bb$ relation. We uniformly choose 14 different directions to look at the sample, all of which are about $z = 0.1$ away from the sample center. The distance is translated by the cosmological parameters used in the Bolshoi simulation. As shown in Fig. 10, we find that redshift distortion has little effects on the $S - bb$ relation. Because the $S - bb$ relation shows mainly the topological information of the cosmic web, it is reasonable that it is not sensitive to local distortions.

Our observation data comes from SDSS DR12 [48]. We take the North Galactic Cap (NGC) as our galaxy zoo

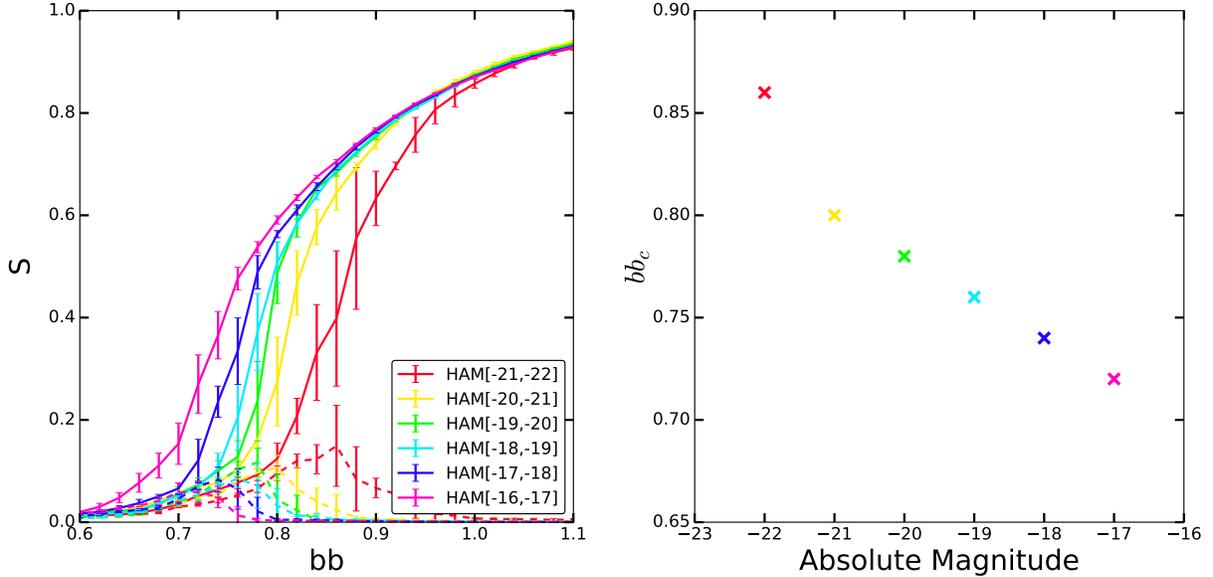


FIG. 7. The $S - bb$ relations of HAM mock samples with different luminosity range are shown here. From left to right, the lines are successively $S1 - bb$ relation and the corresponding $S2 - bb$ relation of HAM Mock samples of $-17 < AM < -16$, $-18 < AM < -17$, $-19 < AM < -18$, $-20 < AM < -19$, $-21 < AM < -20$, $-22 < AM < -21$. We find that the lower luminosity samples' transition bb is smaller than those of the high luminosity samples.

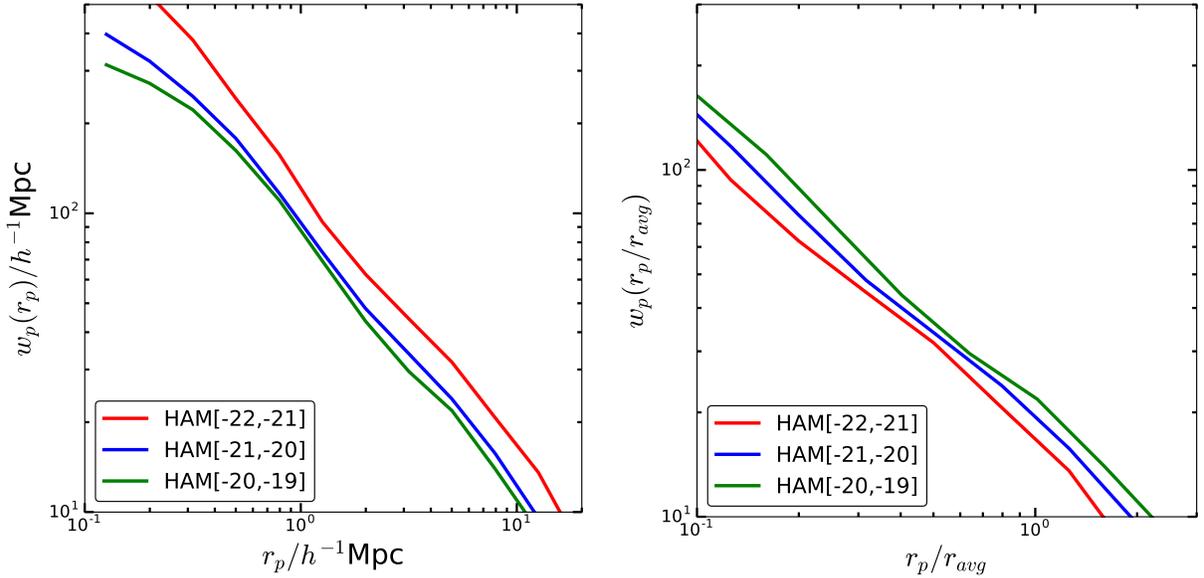


FIG. 8. Comparison between 2-point correlation functions normalized (right) and unnormalized (left) by the point average distance. Lower luminosity samples have higher normalized 2-point correlation function, but they have lower unnormalized 2-point correlation functions, which is well-known. Note that r_p is the projected distance between a pair and r_{avg} is the average distance of the points.

region because of the continuous, large coverage of the NGC. Thus we have the spectrum of galaxies covering huge volume in the red-shift space in NGC. We obtain the data of galaxies in the range of $130 < RA < 230$, $10 < Dec < 50$, and $0.07 < z < 0.14$, including their coordinates, spectral redshifts and Petrosian magnitudes in the r band. The volume of this region is quite similar with the volume of the Bolshoi simulation. Then we use the same cosmological parameters as the Bolshoi simulation to calculate the luminosity distances of the galaxies using their spectral redshifts. We also calculate the Petrosian absolute magnitudes of the galaxies. But we didn't correct their absolute magnitudes.

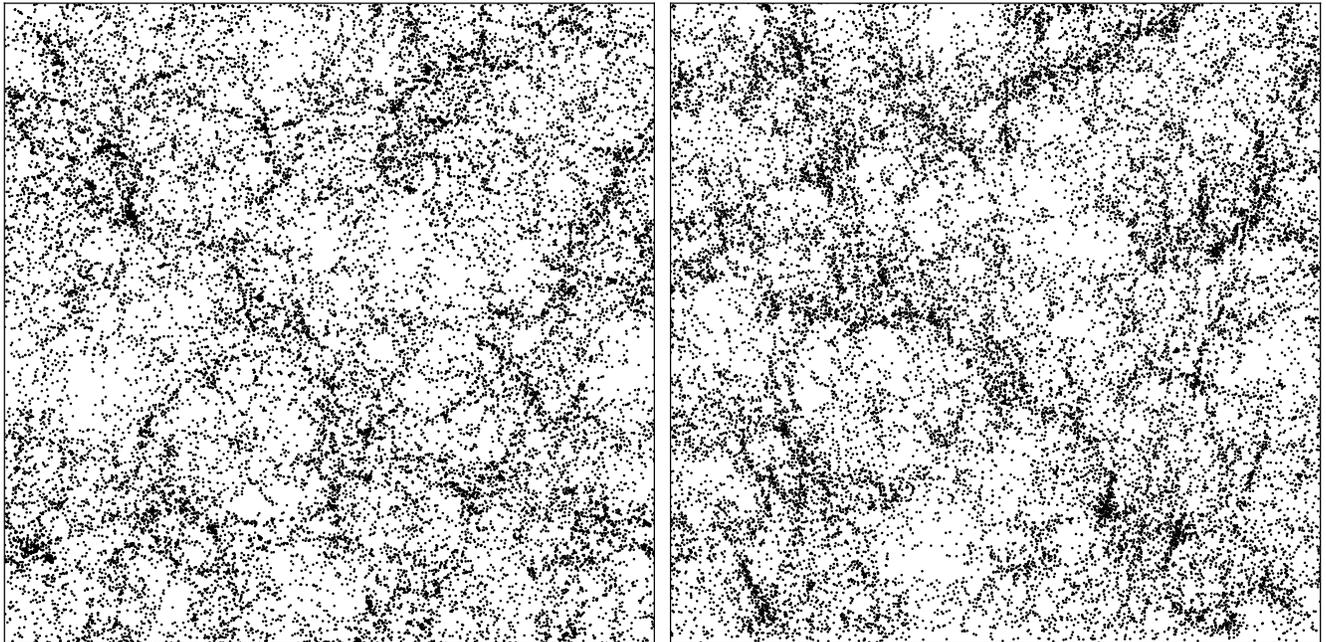


FIG. 9. A slice of galaxy distribution in real space (on the left) and the same slice in redshift space (on the right) as an illustration to show the effect of redshift distortion. The slice is $100h^{-1}\text{Mpc}$ in thickness, $250h^{-1}\text{Mpc}$ in length and width, taken from the mock catalogues we generated. These galaxies' absolute magnitudes are between -20 and -21 .

So the magnitudes of the galaxies are in fact in the $r^{0.1}$ band as described in [55], since the Luminosity Function we use to generate the HAM mock galaxy catalogue is also in the $r^{0.1}$ band. The K-correction for this sample is also not important because we have chosen the red shift range to be $0.07 < z < 0.14$. According to the discussion in [60], the K-correction for $r^{0.1}$ is no more than 0.1. So we just ignore K-correction for our observation sample. Now we have generated an observation galaxy catalogue with 3-D space positions and reasonable absolute magnitudes. The volume covered by the sample is also similar to the volume covered by the HAM mock sample.

Before SDSS DR12, there was no such large continuous angular coverage as well as deep redshift coverage. The public access to the SDSS DR12 allows us to play with a unprecedented, continuous, large scale, 3-D galaxy zoo. It is a good time to search for unusual statistical measurement of the large scale structures, not just the percolation analysis we proposed here but also many other ways of viewing the structures as a whole web.

However, we have to point out that the observational sample is in redshift space while the HAM mock sample is in real space. So the direct comparison of these two samples is not fair. We must take redshift distortion into consideration. Fortunately, our HAM mock sample also contains information of peculiar velocity, which means we can easily transfer the HAM mock sample into the redshift space.

We also need to consider the incompleteness of the observational sample. As shown in the DR12 paper [48], that is mainly because of fiber collision. We can also simulate the incompleteness by excluding those haloes in the simulation that looks closer than $62''$ projected on the 2D plane due to fiber collision, which will exclude $\sim 5\%$ halos mainly in high density region. From Fig. 11 we can see that, this incompleteness has negligible effect for the $S - bb$ relation. Therefore, the $S - bb$ relation is very robust against the observational selection bias introduced by fiber collision. We also find that, changing from the adopted cosmological parameters in both Bolshoi and MDR1 simulations to Planck cosmology, the $S - bb$ relation will not be changed significantly. However, due to the limited boxsize of Bolshoi simulation, the difference between the $S - bb$ relation of MDR1 and Bolshoi simulations is observable. That will cause the misidentification of the transition threshold by 0.04, but for $bb > 0.8$, the $S - bb$ relation difference is negligible. Therefore, using the HAM mock sample built from the Bolshoi simulation to compare with observation is fair, but we need to be aware of a small misidentification of the transition threshold, because of the limited size of the simulation.

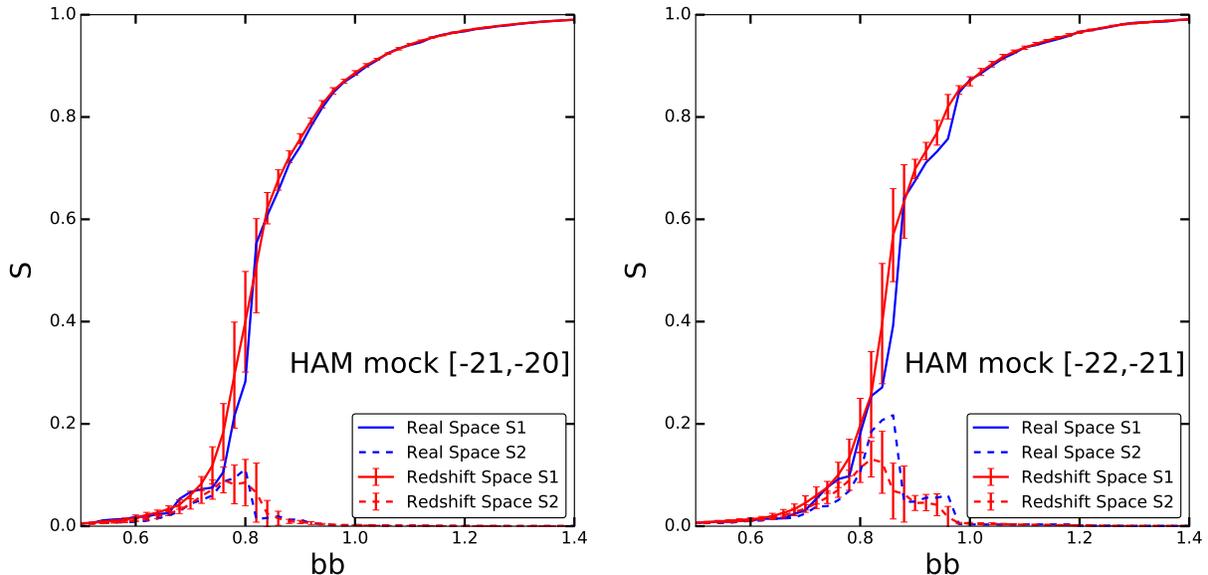


FIG. 10. The $S - bb$ relations of the HAM mock samples in real space (curve without error bars) and in redshift space (curves with error bars). The error bars are given by the standard variance of looking at the same sample from 14 different directions. The center of the sample is at redshift $z = 0.1$. Clearly, the $S - bb$ relation is not sensitive to redshift distortion.

C. Comparison with SDSS DR12

Here we are comparing the observational galaxy sample and the HAM mock galaxy sample in two luminosity ranges $-22 < M < -21$ and $-21 < M < -20$, as shown in Fig. 12. We choose these two ranges because, firstly, we need generally enough number of galaxies to reduce the Poisson error and secondly, we require the galaxies to be detectable even at red shift $z = 0.14$ to get rid of Malmquist bias. Interestingly, although the projected 2-point correlation functions of the HAM mock sample and observed sample are quite similar, their $S - bb$ relations are different by 2-3 σ , particularly the transition point in the $S_2 - bb$ relation. According to our PC CET, this is most probably because the HAM mock sample doesn't reproduce the high order correlation functions well enough. This is also supported by a recent work [62] comparing Illustris simulation and the data sample generated with the same two-point correlation function. The measured $S - bb$ relations (called 'giant component-linking length' relation), for the two samples are significantly different. The transition threshold is also smaller for the real simulation sample.

IV. CONCLUSION AND DISCUSSION

We have proposed the $S - bb$ relation as a branch of percolation analysis of the cosmic web. We have shown that the $S - bb$ relation and its transition threshold reveal physical information of the cosmic web. It also has some advantages compared to the standard 2-point (n-point) correlation function:

1. The $S - bb$ relation is based on the Friend-of-Friend (FoF) algorithm, which is easy to implement and costs very little computational resources.
2. Our theoretical model, PC CET, shows that the $S - bb$ relation contains information of all orders of correlation functions.
3. The $S - bb$ relation is robust against redshift distortion.
4. The $S - bb$ relation is robust against incompleteness, which is very useful in observation. We have shown that 90% completeness is quite acceptable and this has been reached in SDSS.
5. The $S - bb$ relation gives information on the bias of the correlation function.

However, it also has some limitations:

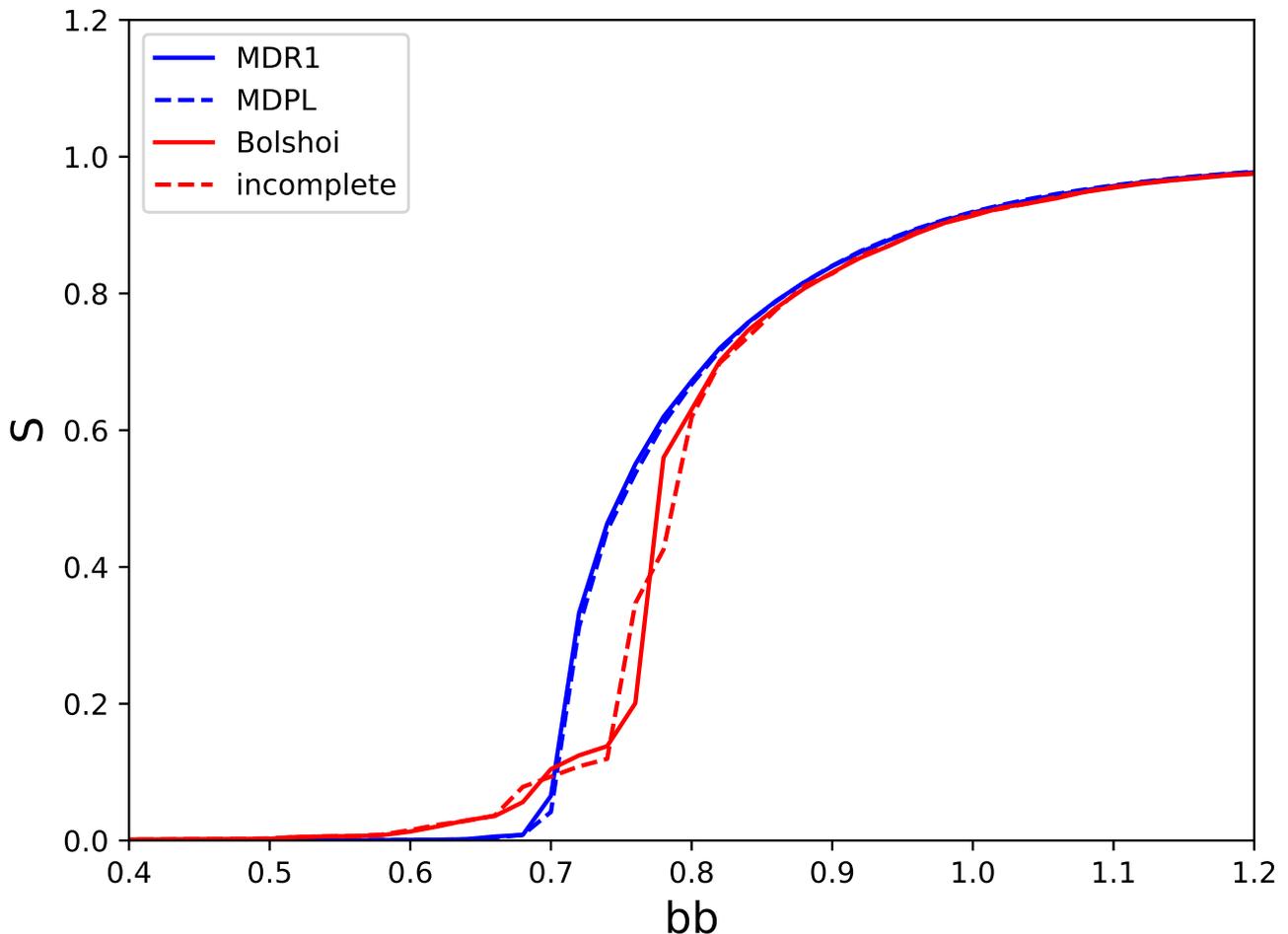


FIG. 11. $S_1 - bb$ relation or simply $S - bb$ relation measured from the halo catalogue with halo mass $> 10^{12} M_\odot$ built from MDR1 simulation [42](WMAP5 cosmology), MDPL simulation [61](Planck cosmology) and Bolshoi simulation [53](WMAP5 cosmology). The red dashed line is measured from the incomplete sample generated from the Bolshoi sample considering fiber collision. The different cosmology between MDR1, Bolshoi and MDPL, and the incompleteness due to fiber collision affect the $S - bb$ relation little. However, the larger boxsizes of the MDR1 and the MDPL simulations introduce sharper transition in the $S - bb$ relation comparing to the Bolshoi simulation.

1. We need large sample of galaxies to look at the $S - bb$ relation, with at least $100h^{-1}\text{Mpc}$ box side length.
2. We need large continuous region of observation to study the $S - bb$ relation, especially for 3-D analysis.
3. It is not easy to extract every order of the correlation functions from the $S - bb$ relation.

Cellular or clumpy? That's not the question any more. It is well known that the distribution of galaxies is cellular. How cellular is the cosmic web? That's the question. We suggest the $S - bb$ relation as a new measurement, which is applicable to a wide range of data types from simulation particles to galaxy catalogue, fast to compute (as fast as 2-point correlation function measurement), robust (against redshift distortion and incompleteness), and it contains information of all orders of correlation functions. In the framework of PCCET, we have good understanding of the $S - bb$ relation. We have compared the $S - bb$ relation of the HAM mock galaxy catalogue with SDSS DR12 galaxy catalogue. A significant difference was found, though these two samples had similar projected 2-point correlation functions. The percolation analysis with discrete points can help us understand the cosmic web. At the minimal, it provides a quick check of whether our mock sample really represents the real distribution or not.

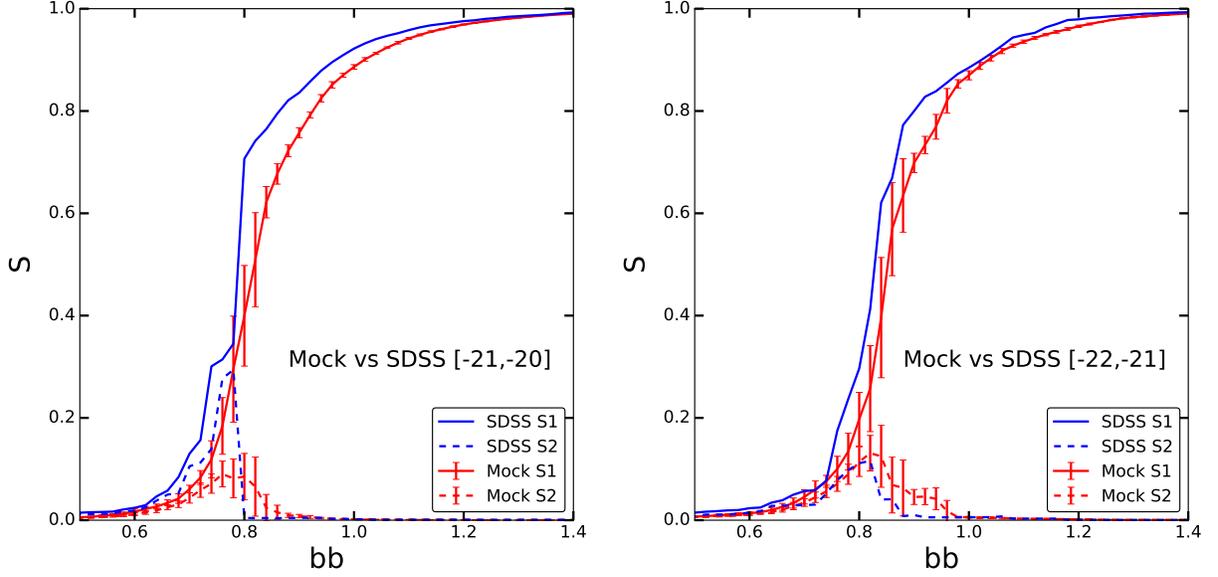


FIG. 12. Comparison of the $S - bb$ relations from SDSS DR12 galaxy sample (blue curves) and the HAM mock sample (red curves). The corresponding $S_2 - bb$ relations are plotted in dashed curves in order to show the transition thresholds. The $S - bb$ relation is different for the mock sample and observation, which is hard to be explained by redshift distortion, incompleteness of observation or different cosmological parameters (Planck and WMAP5).

V. APPENDIX

A. Proof for PCET

The series in Eq. 3 is known q-Pochhammer function or Euler function in mathematics. The q-Pochhammer symbol $(a; q)_n$ is defined as

$$\begin{aligned} (a; q)_n &= \prod_{k=0}^{n-1} (1 - aq^k) \\ &= (1 - a)(1 - aq)(1 - aq^2) \cdots (1 - aq^{n-1}). \end{aligned} \quad (6)$$

In particular, the Euler function $\phi(q)$ is defined as the infinite product

$$\phi(q) = (q; q)_\infty, \quad (7)$$

$$q = 1 - p. \quad (8)$$

The $S_n(p)$ can be expressed in terms of the q-Pochhammer symbols

$$S_n(p) = \frac{1}{n} \sum_{k=0}^{n-1} (q; q)_k, \quad (9)$$

where

$$0 < q < 1. \quad (10)$$

Since $0 < q < 1$, it is not difficult to realise the fact that

$$1 > (q; q)_1 > (q; q)_2 > \cdots > (q; q)_n > 0. \quad (11)$$

Therefore

$$S_n(q) > \frac{1}{n} n (q; q)_{n-1} = (q; q)_{n-1}. \quad (12)$$

On the other hand,

$$S_n(q) = \frac{1}{n} \sum_{k=0}^{n-1} (q; q)_k = \frac{1}{n} \sum_{k=0}^{m-1} (q; q)_k + \frac{1}{n} \sum_{k=m}^{n-1} (q; q)_k. \quad (13)$$

Here we introduce an integer m such that

$$m = INT(\sqrt{n}), \quad (14)$$

where $INT(x)$ is defined as

$$INT(x) = \max\{n \in Z | n \leq x\}. \quad (15)$$

Since as n increases, $(q; q)_n$ decreases, and we have

$$\begin{aligned} S_n(q) &= \frac{1}{n} \sum_{k=0}^{m-1} (q; q)_k + \frac{1}{n} \sum_{k=m}^{n-1} (q; q)_k \\ &< \frac{m}{n} (q; q)_0 + \frac{n-m}{n} (q; q)_m \\ &= \frac{m}{n} + \frac{n-m}{n} (q; q)_m. \end{aligned} \quad (16)$$

In summary,

$$(q; q)_{n-1} < S(n) < \frac{m}{n} + \frac{n-m}{n} (q; q)_m. \quad (17)$$

m as a function of n has the following asymptotic properties

$$\begin{aligned} \lim_{n \rightarrow \infty} m &= \infty, \\ \lim_{n \rightarrow \infty} \frac{m}{n} &= 0. \end{aligned} \quad (18)$$

For simplicity, we denote

$$L_n(q) = (q; q)_{n-1}, \quad U_n(q) = \frac{m}{n} + \frac{n-m}{n} (q; q)_m. \quad (19)$$

Then

$$\begin{aligned} \lim_{n \rightarrow \infty} L_n(q) &= (q; q)_\infty = \phi(q), \\ \lim_{n \rightarrow \infty} U_n(q) &= (q; q)_\infty = \phi(q). \end{aligned} \quad (20)$$

Since we always have

$$L_n(q) < S_n(q) < U_n(q), \quad (21)$$

according to the Squeeze Theorem, we concluded that

$$\lim_{n \rightarrow \infty} S_n(q) = \phi(q). \quad (22)$$

In the form of a step function, we find the limit of $S_n(p)$ to be:

$$\lim_{n \rightarrow \infty} S_n(p) = \begin{cases} 0 & \text{if } p = 0, \\ \phi(1-p) & \text{if } 0 < p < 1, \\ 1 & \text{if } p = 1. \end{cases} \quad (23)$$

ACKNOWLEDGEMENTS

J. Zhang thanks Prof. Y. Jing for giving quite useful advices, L. Yang and Z. Li for helping with measuring 2-point correlation function, Dylan for useful discussion, S. Liao for useful discussion and advice throughout this project, J. Chen for his great help in studying the PCCET mathematically in depth.

The CosmoSim database used in this paper is a service by the Leibniz-Institute for Astrophysics Potsdam (AIP). The MultiDark database was developed in cooperation with the Spanish MultiDark Consolider Project CSD2009-00064.

The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) and the Partnership for Advanced Supercomputing in Europe (PRACE, www.prace-ri.eu) for funding the MultiDark simulation project by providing computing time on the GCS Supercomputer SuperMUC at Leibniz Supercomputing Centre (LRZ, www.lrz.de).

The Bolshoi simulations have been performed within the Bolshoi project of the University of California High-Performance AstroComputing Center (UC-HiPACC) and were run at the NASA Ames Research Center.

-
- [1] M. Davis, J. Huchra, D. W. Latham, and J. Tonry, *The Astrophysical Journal* **253**, 423 (1982).
 [2] M. J. Geller and J. P. Huchra, *Science (New York, N.Y.)* **246**, 897 (1989), jID: 0404511; ppublish.
 [3] Zel'dovich, *Soviet astronomy letters* **8** (1982).
 [4] I. B. Zeldovich, J. Einasto, and S. Shandarin, *Nature* **300**, 407 (1982).
 [5] S. Shandarin, *Soviet Astronomy Letters* **9**, 104 (1983).
 [6] J. Einasto, A. Klypin, E. Saar, and S. Shandarin, *Monthly Notices of the Royal Astronomical Society* **206**, 529 (1984).
 [7] A. Dekel and M. West, *The Astrophysical Journal* **288**, 411 (1985).
 [8] A. Klypin and S. F. Shandarin, *The Astrophysical Journal* **413**, 48 (1993).
 [9] C. Yess and S. F. Shandarin, arXiv preprint astro-ph/9509052 (1995).
 [10] B. Sathyaprakash, V. Sahni, and S. F. Shandarin, *The Astrophysical Journal Letters* **462**, L5 (1996).
 [11] S. Colombi, D. Pogosyan, and T. Souradeep, *Physical Review Letters* **85**, 5515 (2000).
 [12] S. F. Shandarin, J. V. Sheth, and V. Sahni, *Monthly Notices of the Royal Astronomical Society* **353**, 162 (2004).
 [13] S. Shandarin, H. A. Feldman, K. Heitmann, and S. Habib, arXiv preprint astro-ph/0509858 (2005).
 [14] S. F. Shandarin, *Journal of Cosmology and Astroparticle Physics* **2009**, 031 (2009).
 [15] H. Yan and Z. Fan, *The Astrophysical Journal* **730**, 33 (2011).
 [16] S. Shandarin, S. Habib, and K. Heitmann, *Physical Review D* **85**, 083005 (2012).
 [17] G. Brner and H. Mo, *Astronomy and Astrophysics* **224**, 1 (1989).
 [18] C. Yess, S. F. Shandarin, and K. B. Fisher, *The Astrophysical Journal* **474**, 553 (1997).
 [19] B. Pandey and S. Bharadwaj, *Monthly Notices of the Royal Astronomical Society* **357**, 1068 (2005).
 [20] J. ForeroRomero, Y. Hoffman, S. Gottlber, A. Klypin, and G. Yepes, *Monthly Notices of the Royal Astronomical Society* **396**, 1815 (2009).
 [21] M. Boylan-Kolchin, V. Springel, S. D. White, A. Jenkins, and G. Lemson, *Monthly Notices of the Royal Astronomical Society* **398**, 1150 (2009).
 [22] J. R. Bond, L. Kofman, and D. Pogosyan, (1996).
 [23] I. Kayo, Y. Suto, R. C. Nichol, J. Pan, I. Szapudi, A. J. Connolly, J. Gardner, B. Jain, G. Kulkarni, and T. Matsubara, *Publications of the Astronomical Society of Japan* **56**, 415 (2004).
 [24] E. Gaztaaga, P. Norberg, C. Baugh, and D. Croton, *Monthly Notices of the Royal Astronomical Society* **364**, 620 (2005).
 [25] G. V. Kulkarni, R. C. Nichol, R. K. Sheth, H.-J. Seo, D. J. Eisenstein, and A. Gray, *Monthly Notices of the Royal Astronomical Society* **378**, 1196 (2007).
 [26] H. Gil-Marn, J. Norea, L. Verde, W. J. Percival, C. Wagner, M. Manera, and D. P. Schneider, *Monthly Notices of the Royal Astronomical Society* **451**, 539 (2015).
 [27] J. Fry, *The Astrophysical Journal* **277**, L5 (1984).
 [28] C. Hikage, J. Schmalzing, T. Buchert, Y. Suto, I. Kayo, A. Taruya, M. S. Vogeley, F. Hoyle, J. R. Gott, and J. Brinkmann, *Publications of the Astronomical Society of Japan* **55**, 911 (2003).
 [29] M. Calvo, S. F. Shandarin, and A. Szalay, in *Voronoi Diagrams in Science and Engineering (ISVD), 2010 International Symposium on* (IEEE, 2010) pp. 235–243.
 [30] J. M. Colberg, *Monthly Notices of the Royal Astronomical Society* **375**, 337 (2007).
 [31] C. Park, Y.-Y. Choi, M. S. Vogeley, J. R. G. III, J. Kim, C. Hikage, T. Matsubara, M.-G. Park, Y. Suto, and D. H. Weinberg, *The Astrophysical Journal* **633**, 11 (2005).
 [32] B. Robertson, J. S. Bullock, T. J. Cox, T. D. Matteo, L. Hernquist, V. Springel, and N. Yoshida, *The Astrophysical Journal* **645**, 986 (2006).
 [33] G. Kauffmann, S. D. White, and B. Guiderdoni, *Monthly Notices of the Royal Astronomical Society* **264**, 201 (1993).
 [34] Q. Guo, S. White, C. Li, and M. Boylan-Kolchin, *Monthly Notices of the Royal Astronomical Society* **404**, 1111 (2010).
 [35] M. G. Abadi, J. F. Navarro, M. Steinmetz, and V. R. Eke, *The Astrophysical Journal* **591**, 499 (2003).
 [36] S. Trujillo-Gomez, A. Klypin, J. Primack, and A. J. Romanowsky, *The Astrophysical Journal* **742**, 16 (2011).

- [37] A. Klypin, F. Prada, G. Yepes, S. He, and S. Gottlber, *Monthly Notices of the Royal Astronomical Society* **447**, 3693 (2015).
- [38] A. P. Hearin, A. R. Zentner, A. A. Berlind, and J. A. Newman, *Monthly Notices of the Royal Astronomical Society* , stt755 (2013).
- [39] S. Shandarin, S. Habib, and K. Heitmann, *Physical Review D* **81**, 103006 (2010).
- [40] J. Huchra and M. Geller, *The Astrophysical Journal* **257**, 423 (1982).
- [41] V. Springel, *Monthly Notices of the Royal Astronomical Society* **364**, 1105 (2005).
- [42] F. Prada, A. A. Klypin, A. J. Cuesta, J. E. Betancort-Rijo, and J. Primack, *Monthly Notices of the Royal Astronomical Society* **423**, 3018 (2012), arXiv:1104.5130.
- [43] M. Crocce, S. Pueblas, and R. Scoccimarro, *Monthly Notices of the Royal Astronomical Society* **373**, 369 (2006).
- [44] P. Bode, J. P. Ostriker, and N. Turok, *The Astrophysical Journal* **556**, 93 (2001).
- [45] D. J. Eisenstein and W. Hu, *The Astrophysical Journal* **496**, 605 (1998).
- [46] B. L’Huillier, *Astrophysics Source Code Library* (2014).
- [47] Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. J. Banday, R. B. Barreiro, J. G. Bartlett, and et al., *Astronomy and Astrophysics* **594**, A13 (2016), arXiv:1502.01589.
- [48] S. Alam, F. D. Albareti, C. A. Prieto, F. Anders, S. F. Anderson, T. Anderton, B. H. Andrews, E. Armengaud, É. Aubourg, S. Bailey, *et al.*, *The Astrophysical Journal Supplement Series* **219**, 12 (2015).
- [49] M. Sahini and M. Sahimi, *Applications of percolation theory* (CRC Press, 1994).
- [50] S. Kirkpatrick, *Reviews of modern physics* **45**, 574 (1973).
- [51] J. Hoshen and R. Kopelman, *Physical Review B* **14**, 3438 (1976).
- [52] W. Bauhofer and J. Z. Kovacs, *Composites Science and Technology* **69**, 1486 (2009).
- [53] K. Riebe, A. M. Partl, H. Enke, J. ForeroRomero, S. Gottloeber, A. Klypin, G. Lemson, F. Prada, J. R. Primack, and M. Steinmetz, *Astronomische Nachrichten* **334**, 691 (2013).
- [54] A. A. Klypin, S. Trujillo-Gomez, and J. Primack, *The Astrophysical Journal* **740**, 102 (2011).
- [55] A. D. Montero-Dorta and F. Prada, *Monthly Notices of the Royal Astronomical Society* **399**, 1106 (2009).
- [56] S. D. Landy and A. S. Szalay, *The Astrophysical Journal* **412**, 64 (1993).
- [57] I. Zehavi, Z. Zheng, D. H. Weinberg, M. R. Blanton, N. A. Bahcall, A. A. Berlind, J. Brinkmann, J. A. Frieman, J. E. Gunn, R. H. Lupton, *et al.*, *The Astrophysical Journal* **736**, 59 (2011).
- [58] J. M. Bardeen, J. Bond, N. Kaiser, and A. Szalay, *The Astrophysical Journal* **304**, 15 (1986).
- [59] Y. B. Zel’Dovich, *Astronomy and astrophysics* **5**, 84 (1970).
- [60] M. R. Blanton and S. Roweis, *The Astronomical Journal* **133**, 734 (2007).
- [61] A. Klypin, G. Yepes, S. Gottlöber, F. Prada, and S. Heß, *Monthly Notices of the Royal Astronomical Society* **457**, 4340 (2016), arXiv:1411.4001.
- [62] S. Hong, B. C. Coutinho, A. Dey, A.-L. Barabási, M. Vogelsberger, L. Hernquist, and K. Gebhardt, *Monthly Notices of the Royal Astronomical Society* **459**, 2690 (2016), arXiv:1603.02285.