THE NEIGHBORHOOD FUNCTION AND ITS APPLICATION TO IDENTIFYING
LARGE-SCALE STRUCTURE IN THE COMOVING UNIVERSE FRAME

Y.-P. Qin, 1,2,3 L.-Z. Lu, 2,3,4 F.-W. Zhang, 2,4 B.-B. Zhang, 2,4 and J. Zhang 2,4

Received 2007 April 1; accepted 2007 July 12

ABSTRACT

In this paper a relative number density parameter, called the neighborhood function, is introduced so that the crowded nature of the neighborhood of individual sources can be described. With this parameter one can determine the probability of forming a cluster by chance. A method is proposed to identify large-scale structure in the cosmological comoving frame. The scale used to sort out clustered sources is determined by the mean local number density, and the sample adopted is required to have regular borders. The method is applied to a sample drawn from the Two Degree Field survey. We find from the analysis that the probability of forming the resulting large-scale structures by chance is very small, and the phenomenon of clustering is dominant in the local universe. Within a 3σ confidence level, a coherent cluster with a scale as large as 357 h−1 Mpc is identified from the sample. There exist some galaxies which are not affected by the gravitational clusters, and hence are suspected to be at rest in the comoving frame of the universe. Voids are likely volumes within which very crowded sources are not present, and they are likely formed in embryo by fluctuations in the very early epochs of the universe. In addition, we find that large-scale structures are coral-like and are likely made up of smaller structures; sources with large values of the neighborhood function are mainly distributed within the structure of prominent clusters, and they form the framework of the large-scale structure.

Subject headings: cosmology: observations — galaxies: distances and redshifts — galaxies: statistics — large-scale structure of universe — methods: data analysis

Online material: color figures

1. INTRODUCTION

Distributions of galaxies in the sky have been found useful in studying large-scale structures of the universe (Seldner et al. 1977; Gellar & Huchra 1989; Loveday et al. 1992). Since redshift surveys are available from different eras, many investigations have been made to detect large-scale structures in both projected maps and three-dimensional space. Many discoveries of clustering features were made in the 1980s, including voids, galaxy clusters, strings, great attractors, and various great walls (Kirshner et al. 1981; Bahcall & Sonier 1983; Haynes & Giovanelli 1986; Lynden-Bell et al. 1988; Gellar & Huchra 1989). As early as 1989, the CfA Great Wall was found to extend up to 170 h−1 Mpc (Gellar & Huchra 1989), while in recent years the Sloan Great Wall has been observed to extend up to ~400 h−1 Mpc by Gott et al. (2005)6 and Deng et al. (2006).7 The latter is the largest structure in the universe observed so far. Observations also reveal that large-scale structures can be found deep in space where the redshift is high (Shimasaku et al. 2003; Matsuda et al. 2005; Ouchi et al. 2005). There are new opportunities to investigate large-scale structures in the universe with the availability of two new redshift surveys: the 2dF Galaxy Redshift Survey (2dFGRS; Colless et al. 2001) and the Sloan Digital Sky Survey (SDSS; York et al. 2000), which provide large amounts of reliable data.

Large-scale structures in the local universe are of particular importance since theories of structure formation can be directly checked by redshift surveys. It has been predicted that large-scale filamentary or sheetlike overdense regions are preferentially formed in the early universe and later evolve into dense clusters of galaxies, which would be expected in the local universe, and gravitational amplification of the matter density fluctuations that are generated in the early universe is assumed to be responsible for the formation of the clustering structure of the present universe (see Peebles 1982; Blumenthal et al. 1984; Davis et al. 1985, 1992; Governato et al. 1998; Kauffmann et al. 1999; Cen & Ostriker 2000; Benson et al. 2001; Colberg et al. 2005). Filamentary features were found to be real when some redshift surveys were carefully checked (Bhavsar & Ling 1988; Bharadwaj et al. 2000; Bagchi et al. 2002; Ebeling et al. 2004; Pimbblet et al. 2004; Croft et al. 2005; Porter & Raychaudhury 2005). The detection of a large concentration of primeval galaxies at redshift z ~ 3 (Steidel et al. 1998) and the reported discovery of a large-scale coherent filamentary structure of LyC emitters in redshift space at z = 3.1 favor the prediction that galaxies preferentially formed in large-scale filamentary or sheetlike overdensities in the early universe (Matsuda et al. 2005). In agreement with this, protoclusters which have been presumed not to have virialized due to their short cosmological ages (e.g., Venemans 2005) have been identified at as far as z ~ 4 (Venemans et al. 2002; Intema et al. 2006).

The great success of the cold dark matter (CDM) model in recent years makes it a leading model in current studies of cosmology. The charm of the CDM model is its predictive power in the formation of large-scale structures. However, in this model large-scale structures >100 h−1 Mpc are rarely formed, which is challenged by recent observations (see Yoshida et al. 2001; Yoshida 2005). This
suggests that the CDM model predicts a smaller homogeneity scale of the universe than what observations have revealed (note that the large-scale structure in question was found to be several hundred $h^{-1}$ Mpc). In addition, the model faces difficulty in explaining the observed structure on length scales $< 1 h^{-1}$ Mpc as well (see a brief review in Cembranos et al. 2005).

Statistical methods are useful in studying large-scale structures. As mentioned in Martinez & Saar (2002a), most of the statistical analyses of the galaxy distribution proposed so far have been based on second-order methods (correlation functions and power spectra; see also Diggle 1983). Among them, the two-point correlation function (Totsuji & Kihara 1969; Peebles 1974, 1980; Davis & Peebles 1983) and the power spectrum (Fisher et al. 1993; Feldman et al. 1994; Park et al. 1994; Tadros & Efstanthiou 1996) are the tools most often used in both observational and theoretical analyses. The latter is the Fourier transform of the former. Since the two quantities are a Fourier transform pair, complete knowledge of one is equivalent to complete knowledge of the other. But this is not true for their estimators when the samples employed are finite and noisy (Feldman et al. 1994). The two-point correlation function $\xi(r)$, which was first used to measure the strength of galaxy clustering for redshift surveys by Davis & Peebles (1983), describes the excess probability of finding a galaxy in a volume element at a separation $r$ from another randomly chosen galaxy above that for an unclustered distribution. There are various estimators of $\xi(r)$ in the literature. The main difference is their corrections for the edge effect (Martinez & Saar 2002a). The quantity was found to follow a power law, $\xi(r) = (r/r_0)^{-\gamma}$, on small scales ($\sim 10 h^{-1}$ Mpc); for early works, see Totsuji & Kihara 1969; Davis & Geller 1976; Davis & Peebles 1983). In recent investigations, the indices were found to be $\gamma \sim 1.8$ and $r_0 \sim 6.1 h^{-1}$ Mpc for the SDSS data and $\gamma \sim 1.7$ and $r_0 \sim 5.1 h^{-1}$ Mpc for the 2dFGRS data (Zehavi et al. 2002; Hawkins et al. 2003). As pointed out by Yadav et al. (2005), the power-law relation of $\xi(r)$ does not hold on large scales, breaking down at $r > 16 h^{-1}$ Mpc for SDSS and at $r > 20 h^{-1}$ Mpc for 2dFGRS, and thus, it does not violate the homogeneous nature of the universe.

An advantage of the two-point correlation function is its application to galaxy groups to reveal the matter distribution in the universe, since the population of galaxies in a halo is believed to depend on the halo mass (see Benson et al. 2000; Berlind et al. 2003). Since each dark matter halo is expected to give birth to a single group of galaxies, applying the two-point correlation function to different groups of galaxies can reveal the clustering nature of these groups, and this in turn can tell how dark matter is distributed (see, e.g., Jing & Zhang 1988; Merchan et al. 2000; Zandivarez et al. 2003; Padilla et al. 2004).

In measuring the scale of homogeneity of the universe, a power spectrum is applicable (see Einasto & Gramann 1993). As a tool, Martinez et al. (1998) introduced the $K$-function $K(r)$, which is related to the integral of the two-point correlation function. One of their main claims is that the estimators for $K(r)$ are more reliable than the most currently used estimators for $\xi(r)$, which favors the use of $K(r)$ (especially in three-dimensional processes and at large scales) in spite of its somewhat less informative character (see Martinez et al. 1998). In the following, it is found that the concept of the $K$-function is also useful in describing the crowded nature of the neighborhood of individual sources, and hence is able to reveal the density property of a group of sources.

Based on previous methods, we develop in this paper a statistical approach to sorting out clustering sources. Our main concerns include the following: the crowdedness of sources in their neighborhoods are quantified so that one can tell how sources of different crowdedness form a cluster; redshift-dependent densities of samples are taken into account to sort out clusters so that the relevant statistical significance can be determined; and via simulation, probabilities for forming clusters by chance are calculated. The method is applied to the 2dFGRS data set. As a result, some high-resolution maps of large-scale structures derived from the sample are available, and a list of cluster probabilities is obtained. The paper is organized as follows: In § 2 we propose the method, where a statistic called the individual neighborhood function, $s_i(r)$, is introduced to describe the crowded nature of individual sources. In § 3 we describe the selection of the sample, and its redshift distribution is studied. Density distributions of the sample, in terms of $s_i(r)$, are presented in § 4. Clustering probabilities obtained by simulation are studied in § 5. Coherent clusters are identified and shown in § 6. In § 7 we show the structures of some large, prominent, coherent clusters. In § 8 the largest structures sorted out from the whole 2dF data set are shown and discussed. In the last section (§ 9), a summary of the paper and a brief discussion are presented.

2. METHODS AND RELEVANT STATISTICS

2.1. The Clustering Scale

There are two well-known methods for identifying galaxy groups or clusters. Many authors have used the first method, in which projected separation and velocity difference are employed to search for companion sources in order to create galaxy group catalogs (Huchra & Geller 1982; Eke et al. 2004; Diaz et al. 2005; Merchan & Zandivarez 2005). A product of this method is the property associated with the virial theorem (e.g., the velocity dispersion of a set of sources), and it is suitable for identifying virialized groups of galaxies. The second method is used to pick out large-scale structures, where a single scale is applied to all galaxies of a sample to identify their neighboring sources (Einasto et al. 1984; Deng et al. 2006). The main concern of this method is the statistical nature (e.g., the length or density) of the identified clusters. In both approaches, the so-called friends-of-friends (FoF) algorithm, the claim that “any friend of my friend is my friend,” is applied. The latter approach is somewhat suitable for our analysis. In addition to picking out clustering sources, we also need to know the confidence level that a group of sources can be regarded as a cluster or with what probability the cluster can be identified by chance. Here, based on the second method we try to establish a statistical approach to identifying large-scale structures, with which the confidence level of selecting a cluster is able to be obtained.

In order to develop such a method, we need to address several questions: (1) How do we pick out a cluster? (2) At what confidence level could a cluster with a certain number of galaxies and degree of crowding be identified by chance? To find answers to these questions, at least the following requirement should be satisfied: the probability for identifying a group of galaxies as a cluster by chance should be well determined. In doing so, it is essential that all selection effects be removed or avoided. It is plain that if the scale used to find coherent clusters is larger than the area concerned, then all sources within that area must be included in a single coherent cluster; if the scale is as small as possible, then all sources within that area must be separated from each other, and no coherent clusters can be identified. It seems that in identifying a group of clustering sources, one needs to take several steps: determine a criterion clustering scale $r_{ce}$; sort out clustering sources with $r_{ce}$ by applying the FoF algorithm; and calculate the probability for picking out a cluster by chance. In calculating the probability, a Monte Carlo simulation will be used.

In the case of the two-point correlation function, the correlation length $r_0$ is regarded as a characteristic clustering scale.
According to the well-known power-law relation \( \xi(r) = (r/r_0)^{-\gamma} \), the probability of finding another source, relative to any source concerned, at \( r = r_0 \) is twice the probability produced by a random data set. If the sample is large enough, the doubled probability suggests that the density measured at that scale is 2 times the mean density. As mentioned above, in the two recent surveys the mean density. As mentioned above, in the two recent surveys \( r_0 \) was found to range from \(-5.1 \) to \(-6.1 \) h\(^{-1}\) Mpc. A criterion associated with the sorting scale comparable to these scales will be adopted in this paper.

Let \( \rho_0(z) \) be the mean density, expected at redshift \( z \), of a flux-limited but homogeneous data set without the effect of clustering. Similar to the commonly used selection function (see, e.g., Martinez & Saar 2002b; Coil et al. 2004; Padilla et al. 2004), \( \rho_0(z) \) as a function of redshift is determined by the real distribution of the observed galaxies at \( z \), which depends on the evolutionary effect, and is affected by the flux-limited effect, as well as other selection effects such as masks in given fields and fiber collisions in the spectrograph (Martinez & Saar 2002b). As pointed out by Martinez & Saar (2002b), many selection effects are directional, with some being due to the construction of the sample. Absorption by dust in the Milky Way also gives rise to a selection effect that is directional. The best way to take this effect into account is to consider the well-defined maps of the distribution of Galactic dust (Schlegel et al. 1998). When directional selection effects are considered, we should deal with \( \rho_0(z, \theta, \phi) \) instead of \( \rho_0(z) \), where \( \theta \) and \( \phi \) are coordinates of direction. For the sake of simplicity, we consider in this paper only the case of \( \rho_0(z) \), which corresponds to the commonly used radial selection function. The difference between \( \rho_0(z) \) and the radial selection function is that when measuring the former we divide the count of sources within a redshift interval by the real volume confined by that interval.

The criterion clustering scale adopted in this paper is taken as \( r_{ccs} = [2\rho_0(z)]^{-1/3} \). The reason for adopting this criterion clustering scale is that the scale so adopted is generally smaller than, but comparable to, the mentioned typical scales \( r_0 \), which range from \(-5.1 \) to \(-6.1 \) h\(^{-1}\) Mpc in the two recent surveys. The mean of \( r_{ccs} \) in the sample adopted below (sample 2) is 4.2 h\(^{-1}\) Mpc. The number of galaxies in the sample with \( r_{ccs} \) less than 5 h\(^{-1}\) Mpc is 172,265, 76% of the total. That means that sources of clusters will generally be sorted out at scales less than \( r_0 \) (note also that in Deng et al. [2006], \( r \sim 5 \) h\(^{-1}\) Mpc was adopted to sort out clusters, and in Hoyle et al. [2002], \( r = 5 \) h\(^{-1}\) Mpc was used to plot the median density contour). We find that when \( r_{ccs} = 5 \) h\(^{-1}\) Mpc is adopted to sort out clusters, one will get a much larger number of galaxies for the largest coherent cluster with the method proposed below. Adopting \( r_{ccs} = [2\rho_0(z)]^{-1/3} \) will provide us with a conservative result. (Note that when adopting \( r_{ccs} = [4\pi\rho_0(z)^3]^{-1/3} \) one gets a more conservative result, since \( r_{ccs} \) becomes smaller, but this scale will be much smaller compared to \( r_0 \).)

It is known that in the case of the two-point correlation function, \( r_0 \) depends on the clustering property of samples. When adopting \( r_0 \) to identify clusters in a sample, we measure the structure of the sample with a ruler determined from the structure itself. The ruler changes as the structure changes. Unlike \( r_0 \), the criterion clustering scale \( r_{ccs} \) proposed here depends only on the property of the background sample from which the concerned probability will be derived (see below). This quantity is entirely independent of the properties of the adopted sample, and hence is a rather objective ruler.

Due to flux limits, some distant galaxies will be missed, which makes the average distance between the observed distant sources relatively larger. Taking a fixed \( r_{ccs} \) to identify clusters by applying the FoF algorithm, we would probably include all nearby sources in a single large-scale structure if \( r_{ccs} \) were large enough. Meanwhile, for faraway galaxies it might be possible that only a very few of them would be included in a cluster since \( r_{ccs} \) would not be large enough to connect most of these galaxies. This is unfair. For some nearby sources, they might in fact be relatively far away from the local structure but be included as members of the structure since their distances to the structure are shorter than the adopted \( r_{ccs} \); for distant sources, since the average distances are large they have less chance of being identified as members of a cluster according to this fixed \( r_{ccs} \). When we adopt \( r_{ccs} = [2\rho_0(z)]^{-1/3} \) we deal with a varying \( r_{ccs} \), since \( \rho_0(z) \) is a function of redshift. In this case we get a large value of \( r_{ccs} \) for higher redshifts and obtain a smaller value for lower redshifts. As part of our method, a varying \( r_{ccs} \) will be adopted to identify clusters (see below).

### 2.2. Radial Selection Function

The result of a simulation would be acceptable if the background sample so created had no bias or the bias were insignificant. It is essential that all selection effects and evolutionary effects be removed or avoided. There is a straightforward technique to avoid these effects. Assume that within a spherical space centered at the observer, there are \( N_0 \) galaxies in total, which form various structures due to dynamics and other known or unknown factors. When no clustering mechanisms are at work, these sources are expected to be homogeneously distributed within the area, and they constitute a global homogeneous background sample. In fact, what one can observe at present are galaxies of different cosmological ages. Due to evolutionary effects, number densities might evolve with redshift, and then the total number \( N \) of galaxies whose photons could reach us at present would be different from \( N_0 \) (e.g., the formation and merging of galaxies might occur somewhere at some time). In this situation the distribution of the observable background sample would only be homogeneous locally.

Let us consider a survey over the whole sky. Due to the flux limits and other selection effects, only \( N' (N' < N) \) galaxies could be observed in both the real and background samples, which we call the whole observed real sample and the whole observed background sample, respectively. (Note that these two samples have the same number of galaxies, but the galaxies are assumed to be distributed randomly in the latter sample.) It is obvious that the mean number densities of the whole observed real sample and the whole observed background sample would be the same at different redshifts, from which we can deduce a mean number density law with respect to redshift. Now we perform a survey over a limited area of the sky. According to the isotropic nature of the universe, it is expected that the mean number densities of this smaller observed real sample and the corresponding observed background sample would be the same, as well at different redshifts. Keeping this in mind, a background sample created by the simulation following this mean number density law would have no bias. The statistical significance derived from this sample would be safely acceptable. Note that selection effects would differ from survey to survey due to different techniques and instruments. The proposed simulation method is simple due to the fact that it does not depend on the details of selection effects (in fact, all selection effects have been taken into account).

Besides the selection effects, there exists a boundary effect for most samples. It is desirable for problems arising from this effect to be taken care of. The most effective way of solving this problem is to consider the real volume involved. When estimating \( \xi(r) \), the count of sources in the adopted sample lying inside the shell \( [r - \Delta r/2, r + \Delta r/2] \) relative to a point is divided by the volume of the intersection of that shell with the whole sample volume.
density of the background sample expected at the redshift corresponding to $r$. Concerning the crowded nature of the neighborhood (the neighborhood density) of a source, let us introduce an individual neighborhood function $\kappa_i(r)$ for object $i$ to describe the number density of the sample within scale $r$ relative to the object, which is defined by

$$\kappa_i(r) \equiv \frac{\int_{V(r)} \rho(r) \, dV}{\rho_0(z_i)V_i(r)} - 1,$$

where $z_i$ is the redshift of object $i$ and $V_i(r)$ is the volume of the intersection of the sphere with radius $r$, centered at object $i$, with the whole volume of the adopted sample (here $r$ does not represent $|r|$). The mean of $\kappa_i(r)$ of a sample (called the neighborhood function of the sample),

$$\kappa(r) \equiv \frac{1}{N} \sum_{i=1}^{N} \kappa_i(r),$$

is one of its statistical properties (a spatial distribution property), reflecting the mean relative number density of the sample measured within the mentioned scale (note that when adopting $r = r_{\text{cess}}$, $r$ will differ from source to source). For a subset of the sample (namely, a group of sources of the sample) we also use the mean of $\kappa_i(r)$ to describe its spatial distribution property (its density nature), and this quantity is also denoted by $\kappa(r)$. The only difference is that, in equation (3), $N$ will be replaced by $N'$, where $N'$ is the number of galaxies of the subset. It is obvious that the larger the $\kappa(r)$, the more crowded will be the sample or the group. In terms of $\eta(r)$ the two quantities can be expressed by

$$\kappa_i(r) = \frac{\int_{V(r)} \eta(r) \, dV}{V_i(r)},$$

$$\kappa(r) = \frac{1}{N} \sum_{i=1}^{N} \frac{\int_{V(r)} \eta(r) \, dV}{V_i(r)}.$$

[Note that when different weights for the particles are adopted and taken into account, one might deal with other definitions of $\kappa_i(r)$ and $\kappa(r)$. In that situation, one might wish to modify definition (1) so that the forms of other equations are maintained.]

In the case of the two-point correlation function, $\xi(r) > 0$ suggests that, for any given source, the probability of finding another source within a distance $r$ is larger than the probability produced by a random data set, and the former is $1 + \xi(r)$ times the latter. In the case of an individual neighborhood function, the probability of finding a source within the scale $r$ relative to position $r_i$ is proportional to the integral of the number density $\rho(r)$ over the volume confined by $|r - r_i| < r$ for the adopted sample; it is $1 + \kappa_i(r)$ times that for the background sample, which is created randomly according to $\rho(z)$, where $\kappa_i(r) > 0$ gives a larger probability for the sample than for the background sample. In the case of the neighborhood function, the probability of finding a source within the scale $r$ is $1 + \kappa(r)$ times that for the background sample.

---

8 For example, when one assumes a real density $\rho_0(z)$ but deals with a sample for which some galaxies are missed, one might wish to introduce a weight to account for the missing sources. In this way, $\bar{\rho}(r)$ would be replaced by $\bar{\rho}(r) w(z)$, where $w(z)$ is the weight which is assumed to be known, and then eq. (1) is modified. Absorbing $w(z)$ into $\rho_0(z)$, $\rho_0(z) = \rho_0(z)/w(z)$, and replacing $\rho_i(z)$ with $\bar{\rho}(z)$, one will get the same form of equations for $\kappa_i(r)$ and $\kappa(r)$. 
In practice, $\kappa_s(r)$ and $\kappa(r)$ can be computed by

$$\kappa_s(r) = \frac{1}{\rho_0(z)} \Delta V_s(r) \sum_j \int n_j dV - 1,$$

and

$$\kappa(r) = \frac{1}{N} \sum_i \frac{1}{\rho_0(z)} \Delta V(r) \sum_j \int n_j dV - 1,$$

where $n_j$ is the $\delta$-function giving the position of particle $j$; $\Delta V_s(r)$ is the portion of the whole sample volume satisfying $|r - r_j| < r$ for any position $r$; the integral is taken over $\Delta V(r)$; and the sum over $j$ includes only particles within $|r_j - r| < r$ in the sample, which yields the number of sources found within $\Delta V(r)$. For a volume-limited sample, $\Delta V(r)$ would differ from source to source due to the different positions of the objects relative to the borders of the sample volume, even if $r$ were the same.

One can check that, when ignoring the variation of $\rho_0(z)$ with respect to redshift, the quantity $1 + \kappa(r)$ would be similar to the Ripley $K$-function, which is an integral of $1 + \xi(r)$ over a spherical volume with radius $r$ (Ripley 1981; Martinez et al. 1998). There is no direct relation between the $\kappa(r)$ estimator and the commonly used $\xi(r)$ estimators, since we calculate $\kappa(r)$ in a slightly different manner [e.g., no random samples are required to compute $\kappa(r)$]. However, when the difference between $1 + \kappa(r)$ and the Ripley $K$-function is ignored, the differential of $1 + \kappa(r)$ becomes $1 + \xi(r)$. The main difference between the $K$-function, or the two-point correlation function, and the neighborhood function is that with $\kappa_s(r)$ we can tell how crowded the neighborhood of a source is and can even show how sources of different crowdness are distributed in the space and how this distribution is related to the structures observed (see below).

The quantity $\kappa(r)$ could also be closely related to the conditional density $\Gamma^*(r)$ in a sphere (Vasilyev et al. 2006). Applying equation (7), when one takes $\rho_0(z) = 1$ and takes the sum over the whole sample, $1 + \kappa(r)$ becomes $\Gamma^*(r)$, as long as the volume of the sample is large enough so that, for any point in the sample, the sphere of radius $r$ centered at the point is well inside the volume.

In the following analysis, cosmological distances, as well as comoving volumes, will be expressed in terms of the comoving coordinates, and in calculating the latter the standard cosmological model, $(\Omega_M, \Omega_\Lambda) = (0.3, 0.7)$, will be adopted throughout this paper, leaving the Hubble constant parameter $h (H_0 = 100 h \text{ km s}^{-1} \text{Mpc}^{-1})$ serving as a unit. The analysis will be performed in redshift space. That is, when calculating the comoving coordinates of galaxies from samples drawn from redshift surveys, the peculiar velocity of individual sources will be ignored, and the measured redshift will be taken as the real distance indicator of the object.

### Table 1: Description of Samples

| Name       | Number | Region  | $\rho_0(z)$ | Data  |
|------------|--------|---------|-------------|-------|
| Sample 1   | 59,497 | NGP     | $\rho_1$    | Observation |
| Sample 2   | 226,302| NGP, SGP| $\rho_1$    | Observation |
| Sample 3   | 59,497 | NGP     | $\rho_1$    | Simulation |
| Sample 4   | 59,497 | NGP     | $\rho_2$    | Simulation |
| Sample 5   | 59,497 | NGP     | $\rho_3$    | Simulation |

a For definitions of NGP and SGP, see Colless et al. (2001).

b In this column we list the $\rho_0(z)$ that is used to perform the corresponding simulations, to calculate the corresponding $\kappa$, and/or to determine the corresponding $r_{cc}$, where $\rho_1$ is that directly measured from sample 2, $\rho_2$ is the polynomial function obtained by a fit to $\rho_1$, and $\rho_3 = \rho_0 = 0.00247 \text{ (h}^{-1} \text{ Mpc})^{-3}$ is a constant.

c Taken over 150.0\' $<$ R.A. $<$ 220.0\', $-4.8\' < $ decl. $< 1.0\'$.

d In addition to the NGP and SGP strips, sample 2 covers all other “random” fields of the 2dF survey.

3. SAMPLE SELECTIONS AND REDSHIFT DISTRIBUTIONS

The data studied are taken from the 2dFGRS final data release (Colless et al. 2003). Two main regions of the survey data located in distinct areas of the sky are visible. To calculate $\kappa_s(r)$, as well as $\kappa(r)$, we must know the volume $\Delta V_s$, for each source and for any given $r$. This is achievable if the area concerned is well defined with sharp borders. The computation is also easier if the edges of the region are regularly cut. The following criteria are adopted to select a sample from the survey data set: $150.0'$ $<$ R.A. $(J2000.0) < 220.0'$, $-4.8'$ $<$ decl. $(J2000.0) < 1.0'$, $0.0' < z < 0.3$, and quality $\geq 3$, where “quality” is the redshift quality parameter for the best spectrum (quality = 1–5; reliable redshifts have quality $\geq 3$). We thus get a sample of 59,497 sources (sample 1). Table 1 presents the description of this sample, as well as other samples mentioned below, where column (1) is the name of the sample, columns (2) and (3) denote the size and region of the sample, respectively, column (4) shows the $\rho_0(z)$ that is used to calculate $\kappa$ for the samples and to create the corresponding background samples by simulation (see below), and column (5) denotes the type of data (observation or simulation, with the former being that of a real sample and the latter being that of a background sample). The sky region of sample 1 is illustrated in Figure 1, where the one of the two distinct areas that encloses this sample is also presented.

We notice that there are shortcomings of the 2dF sample. As shown in Cannon et al. (2006), there are fiber collisions within 2dF, where many objects around the edge of a 2\' field are missed (see Fig. 3 in Cannon et al. 2006). This gives rise to a directional incompleteness in the survey. However, the available data are from a set of 2\’ fields such that most regions of the sky inside the survey boundary are covered by several overlapping fields, and the potential incompleteness has been significantly reduced (see Fig. 5 in Colless et al. 2003). In addition, we cut away the edge of the region concerned, and this reduces the directional effects of fiber collisions as well. We presume in this paper that this effect is negligible, and hence can be ignored.

Another shortcoming is the existence of “missed galaxies” (see Pimbblet et al. 2001; Cross et al. 2004). As shown in Cross et al. (2004), the incompleteness of the sample could reach about 14\%, which varies slightly with magnitude (see Fig. 11 of the paper). This shortcoming is a redshift-dependent selection effect. As a function of redshift, it is absorbed into $\rho_0(z)$ (see below).

The third shortcoming is the fact that the $b_j = 19.45$ mag limit is not constant across the entire area of the survey (see Colless et al. 2001; Norberg et al. 2002). This might affect the results of the analysis. We will discuss this effect in the last section of this paper.

Since the volume of the area concerned is well defined, we are now able to estimate the mean density of the background sample, $\rho_0(z)$, from the observational data. As pointed out by Padilla et al. (2004), there are several approaches to doing so. The most straightforward way is to use the luminosity function of galaxies to obtain an estimate of $\rho_0(z)$. This is not adopted in this paper, since any bias from the real luminosity function would cause systematic errors (see Padilla et al. 2004 and the references therein). In the following, let us consider three other approaches. The first one is to estimate $\rho_0(z)$ directly from the observational data. In
this approach we assume that the redshift distribution of the whole 2dF survey data set could serve as a parent population of the background sample. Of course, to match the data of sample 1, the data set serving as the parent population, which covers the whole area of the 2dF survey, should meet these criteria: $0 < z < 0.3$ and quality $\geq 3$. A set of 226,302 sources (sample 2) is obtained, which is almost 4 times the number of galaxies in sample 1. As illustrated below, there are two massive superclusters in 2dF, and these superclusters are comparable to the survey volume (see also Erdogdu et al. 2004; Lahav & Suto 2004; Porter & Raychaudhury 2005; Einasto et al. 2007a). This suggests that 2dF might not be a fair sample of the universe. The reason for assuming the whole 2dF survey data set to be a fair sample of the universe is that it is indeed very large (many statistical analyses have been based on it), and it is the largest available sample of this kind (detected by the same means as the 2dF survey). Under this assumption, we simply measure $\rho_0(z)$ from sample 2 and then apply it to our simulation analysis. The advantage of this approach is that any possible evolutionary and selection effects, known or unknown, will be accounted for. There are two disadvantages of the method. One is due to the measurement uncertainty, which causes an unreal distribution within the uncertainty scale. This is eased by assigning a random value smaller than the uncertainty to the background data when randomly drawing from sample 2 (the following applications show that this technique indeed works; see Fig. 3). The other disadvantage is due to the fact that the 2dF survey does not cover the whole sky. If there exist large clusters which are comparable to the regions concerned (as pointed out above, there are some), the redshift distribution of the presumed parent population will be affected: bumps will be seen in very dense places, and troughs will exist in very sparse spaces. In this situation, some redshift distribution feature due to clustering (Padilla et al. 2004) could be mistaken as a background property, and then the clusters corresponding to this feature would be missed (at least in redshift space this clustering property is ignored). Thus, the number of galaxies in some very large clusters would become smaller. To get a safer number of galaxies, one might prefer this approach.

The second way of estimating $\rho_0(z)$ is to fit parametric forms to the observational data (see Padilla et al. 2004), which uses sample 2 as well. In this approach, we do not directly apply $\rho_0(z)$ measured from sample 2 to create background samples. Instead, we fit it with an empirical curve. As long as the probability obtained from the fit is small enough, the fitting curve will serve as $\rho_0(z)$, and it will be applied to our simulation analysis. The mean density curve so obtained is nothing but a deep smoothing of that directly measured from sample 2. There is a risk in employing this approach. In the process of smoothing, while bumps and troughs caused by clustering will be smoothed, features arising from selection effects will be smoothed as well. In the following analysis, results arising from the first and the second approaches will be presented and compared.

For the sake of comparison, the third approach to estimate $\rho_0(z)$ is also adopted. In this approach, the galaxies observed are assumed to be homogeneously distributed over the whole area of sample 1. Thus, we simply take $\rho_0(z) = \text{const}$ to create background samples. Although the redshift distribution of this kind of background sample is not real, the analysis might reveal some aspects of clustering which were previously unfamiliar (see below). Let us derive $\rho_0(z)$ from sample 2. The count observed within any sky region of the 2dF survey is assumed to be proportional to the corresponding sky area according to the isotropic property of the universe. Let $z_i$ be the redshift of source $i$ in sample 2. We calculate the count within the redshift interval $0.99z_i - 1.01z_i$ from sample 2 and calculate the volume $V(\Delta z_i)$ confined by this redshift interval in the area of sample 1. According to the isotropic assumption, the volume confined by this redshift interval in the whole area of sample 2 should be $226,302/59.497$ times $V(\Delta z_i)$. Thus, the count divided by $(226,302/59.497)V(\Delta z_i)$ will be taken as the estimate of $\rho_0(z)$ for this source, which is denoted by $\rho_0(z_i)$. In doing so, when the redshift interval is less than the measurement uncertainty of the survey, the latter will be taken as the width of the redshift interval.

The estimated values of $\rho_0(z_i)$ of sample 2 are shown in Figure 2. The data can be well fitted by the polynomial function

$$
\log \rho_0(z) = -(11.2931 \pm 0.0049) - (20.585 \pm 0.013) \log z
- (15.587 \pm 0.011)(\log z)^2
- (4.9779 \pm 0.0039)(\log z)^3
- (0.52885 \pm 0.00045)(\log z)^4
$$

($P < 0.0001$), where $\rho_0(z)$ is in units of $(h^{-1} \text{Mpc})^{-3}$.

We create a background sample (sample 3) randomly from sample 2 with the first approach. The second background sample (sample 4) is produced with the second approach by simulation.
according to the empirical mean number density function, i.e., the polynomial function. The third background sample (sample 5) is yielded by simulation under the assumption that the galaxies are homogeneously distributed, where the mean number density is obtained by dividing 59,497 by the whole volume of sample 1, which gives \( \rho_0 = 0.00247 \, (h^{-1} \text{ Mpc})^{-3} \). The numbers of galaxies in all these background samples are the same as in sample 1, \( N = 59,497 \).

The redshift distributions of samples 1–5 are shown in Figure 3. Figure 3a shows that there indeed exist some bumps and troughs in both samples 1 and 2. Comparing with the other panels, we suspect that this phenomenon is unlikely to be caused by randomness. Instead, it might be due to intrinsic distributions such as large-scale structures (see the discussion above in this section and Fig. 2) or selection effects. In addition, we find some significant features in sample 1 which are much less prominent in sample 2. This must be due to a local property of the redshift distribution and might be a consequence of the existence of local clustering (see Figs. 17, 20, and 21). Figure 3b clearly demonstrates that samples created with the first approach well follow the redshift distribution of sample 2. In addition, it reveals that the noise in the observational samples is largely caused by the measurement uncertainty (since adding an additional random value smaller than the uncertainty significantly reduces the fluctuation observed, as Fig. 3b shows). From Figure 3c we observe that when ignoring the fluctuations, which might be caused by local clustering, the polynomial function could indeed well represent the real distribution of the observational samples. Figure 3d illustrates that, as expected, the sources observed are not homogeneously distributed at all (the flux limit and the evolutionary effect would be the main factors accounting for the nonhomogeneous distribution observed in the survey data).

Although the spatial distributions of the sources in background samples 3 and 4 differ significantly from that of the real homogeneous sample, sample 5, a distribution in agreement with the former two is still regarded as homogeneous in the following analysis. This is because background samples 3 and 4 suffer from flux limits and the evolutionary effect, as well as other known or unknown selection effects, and it is these effects that lead to the apparent inhomogeneous distribution. When all galaxies of the same cosmological age are available, the spatial distribution of the background sample sources, which are created randomly, must be homogeneous due to principles of cosmology. Therefore, in the analysis below, spatial distributions of sources which are in agreement with that of the adopted background sample, or in agreement with that of the parent population, will be regarded as intrinsically homogeneous. Otherwise, they will be considered inhomogeneous.

In Figure 2 we notice a deviation of the fitted polynomial for \( \rho_0(z) \) from the mean number density function measured directly from sample 2. In addition, a deviation of the redshift distribution of the polynomial function from that of the \( \rho_0(z) \) of sample 2 is seen in Figure 3 (see Figs. 3b and 3c). This suggests that the polynomial fit for \( \rho_0(z) \) and the \( \rho_0(z) \) measured from sample 2 are not the same. However, as one will see below, this difference does not cause a major problem in the clustering analysis performed in this paper. Since sample 2 suffers from the clustering effect, as mentioned above and revealed below, it might be possible that samples 3 and 4, which come from sample 2, would be inhomogeneous not only due to the flux limit but also due to clustering. Thus, it needs to be checked if the deviation observed in Figure 2 leads to an observable difference in the homogeneity properties of the two corresponding background samples. We check this by examining their \( K \)-functions, from which the corresponding correlation dimension \( D_2 \) can be derived (see Martinez et al. 1998). The \( K \)-function adopted to investigate this issue is that provided in equation (8) of Martinez et al. (1998), with the standard estimator introduced
by Doguwa & Upton (1989) to account for the boundary effect. As is generally known, regardless of the possible slight bias, this estimator, $K_{est}$, has good properties. In Figure 4 one can find the $K$-functions estimated from samples 3 and 4, and for comparison the $K$-functions from samples 1 and 5 are also presented. As expected, the $K$-function estimated from sample 5 follows well the homogeneity curve, the curve for which $D_2 = 3$. The three other samples are seen not to be homogeneous at all (for all of them, $D_2 < 3$ holds within the maximum scale of the samples). It is interesting that the $K$-functions estimated from samples 1, 3, and 4 follow almost the same trend (in this case, they have almost the same $D_2$). The $K$-function from sample 1 is slightly larger than the $K$-functions of samples 3 and 4. However, the difference between the latter two is not detectable. This suggests that the $K$-function, and hence the derived $D_2$, suffer mainly from the redshift distribution, which is influenced by the flux limit, the possible evolutionary effect, and the spatial distribution of the large-scale structure members (see Fig. 3). The role that the local clustering (sources in sample 1 are strongly clustered while those in samples 3 and 4 are not; see the spatial clustering plots presented in Figs. 6–8) plays in producing the correlation dimension $D_2$ seems very insignificant.

4. DENSITY DISTRIBUTIONS

Here, we use the estimator in equation (6) to calculate the individual neighborhood function of sources in the scale of $r_{ccs}$ (i.e., we take $r = r_{ccs}$), which is determined by $r_{ccs} = [2\rho_0(z)]^{-1/3}$. We apply the directly measured $\rho_0(z)$ from sample 2 to calculate $\kappa_i(r_{ccs})$ for each source of samples 1 and 3. In addition, we determine $\rho_0(z)$ using the polynomial function obtained above, and then with this $\rho_0(z)$ calculate $\kappa_i(r_{ccs})$ for each source of samples 1 and 4. Also, we adopt $\rho_0 = 0.00247 \,(h^{-1} \,\text{Mpc})^{-3}$, the mean number density of sample 1, to calculate $\kappa_i(r_{ccs})$ for each source of samples 1 and 5. In this way we calculate $\kappa_i(r_{ccs})$ for sample 1 with three kinds of $\rho_0(z)$ (and hence with three kinds of $r_{ccs}$) estimated with three approaches.

Distributions of the individual neighborhood function $\kappa_i(r_{ccs})$ of sample 1, calculated with three kinds of $\rho_0(z)$, are presented in Figures 5a, 5c, and 5e, respectively. Meanwhile, the distributions of $\kappa_i(r_{ccs})$ of samples 3, 4, and 5 are displayed in Figures 5b, 5d, and 5f, respectively. One might observe that, as histogram plots, the distributions in Figure 5 look like continua, since the total number involved is very large. However, with the same total number, the distributions in Figure 5 look like histograms. We have checked that this difference is not due to the adopted bins. The distribution plots shown in Figure 5 look like histograms because in some ranges of $\kappa_i(r_{ccs})$ the counts are relatively small, while in the nearby $\kappa_i(r_{ccs})$ ranges the corresponding counts are large (thus, the former counts look like zero). This phenomenon is due to the small scale adopted by us (i.e., $r_{ccs}$). Since the scale is relatively small (see the definition of $r_{ccs}$ presented above), one can only find small numbers of sources within the corresponding volume. Thus, the number difference would lead to an obvious difference in the individual neighborhood function. For example, the volumes confined by $r_{ccs}$ would differ slightly for two closely located sources. If the number of neighborhood galaxies relative to one source were three and that relative to the other were four, then this difference would cause a 30% change (in this situation it would be hard to get 3% change for these two sources). For the background samples, the individual neighborhood function is mainly distributed within

![Figure 5](image-url)

**Fig. 5.—Distributions of the individual neighborhood function for samples 1 (black lines in panels a, c, and e), 3 (black lines in panel b, gray lines in panel a), 4 (black lines in panel d, gray lines in panel c), and 5 (black lines in panel f, gray lines in panel e). The individual neighborhood functions presented in (a) and (b) are calculated with the values of $\rho_0(z)$ directly estimated from sample 2 (the first approach); those in (c) and (d) are calculated with the values of $\rho_0(z)$ determined by the empirical polynomial function (the second approach); and those in (e) and (f) are calculated with the constant mean density $\rho_0 = 0.00247 \,(h^{-1} \,\text{Mpc})^{-3}$ (the third approach).**

...cantly influence the distribution of $\kappa_i(r_{ccs})$ for the adopted real sample, sample 1. (See also the discussion below for the case of spatial distributions.)

One might observe that, as histogram plots, the distributions in Figure 3 look like continua, since the total number involved is very large. However, with the same total number, the distributions in Figure 5 look like histograms. We have checked that this difference is not due to the adopted bins. The distribution plots shown in Figure 5 look like histograms because in some ranges of $\kappa_i(r_{ccs})$ the counts are relatively small, while in the nearby $\kappa_i(r_{ccs})$ ranges the corresponding counts are large (thus, the former counts look like zero). This phenomenon is due to the small scale adopted by us (i.e., $r_{ccs}$). Since the scale is relatively small (see the definition of $r_{ccs}$ presented above), one can only find small numbers of sources within the corresponding volume. Thus, the number difference would lead to an obvious difference in the individual neighborhood function. For example, the volumes confined by $r_{ccs}$ would differ slightly for two closely located sources. If the number of neighborhood galaxies relative to one source were three and that relative to the other were four, then this difference would cause a 30% change (in this situation it would be hard to get 3% change for these two sources). For the background samples, the individual neighborhood function is mainly distributed within
the range of \( \kappa_i(r_{ccs}) < 3 \) (see Figs. 5b, 5d, and 5f), while for sample 1, a large number of sources have \( \kappa_i(r_{ccs}) \) larger than 3 (see Figs. 5a, 5c, and 5e).

Motivated by Figure 1 of Governato et al. (1998), let us define a region within which the initial matter forms a presumed galaxy at a later time (the matter is referred to as the "presumed galaxy matter") and assign the central position of the region as the position of that presumed galaxy. These presumed galaxies are expected to be randomly scattered in space. In terms of statistics, they are considered as unclustered sources. In our analysis, the \( \kappa_i(r_{ccs}) \) distributions of unclustered galaxies are those displayed in Figures 5b, 5d, and 5f. The \( \kappa_i(r_{ccs}) \) distribution of sample 1 is obviously different from them (see Figs. 5a, 5c, and 5e). Revealed in the figure, a large number of sources in sample 1 possess very large values of \( \kappa_i(r_{ccs}) \) (say, \( \kappa_i > 3 \)), which is the majority of unclustered galaxies do not have, suggesting that there is clustering in this sample.

Based on Figure 5 (comparing the right and left panels), we define galaxies with \(-1 \leq \kappa_i < 1, 1 \leq \kappa_i < 3, \) and \( 3 \leq \kappa_i \) as uncrowded, crowded, and very crowded sources, respectively. In other words, sources with \(-1 \leq \kappa_i < 1, 1 \leq \kappa_i < 3, \) and \( 3 \leq \kappa_i \) are regarded to be located in regions with relatively small, large, and very large number densities of galaxies, respectively. They have different relative number density environments. The spatial distributions of sources of different crowdedness for samples 1, 3, 4, and 5 are displayed in Figures 6–8. As expected, one can find from Figures 6a and 7a that very crowded galaxies seem to form the core of large-scale structures, and surrounding the core are less crowded sources (for identifying large-scale structures, see the analysis below). When adopting the mean number density of the sample to evaluate the criterion clustering scale, the number of very crowded sources identified from sample 1 (Fig. 8a) is very large (45,753 sources; 77% of the total number of the sample) due to the smaller value of the density, \( \rho_0 = 0.00247 \ (h^{-1} \ \text{Mpc})^{-3} \), but the region these sources occupy is smaller than in other cases. (In Figs. 6a and 7a, the numbers of very crowded galaxies are 27,312 and 27,476, respectively.) Also as expected, all the uncrowded, crowded, and very crowded sources in the background samples are seen to be homogeneously distributed [relative to the expected observed density \( \mu(z) \)] over the whole sample volume (see panel b in the three figures; a similar issue will be investigated below; see Fig. 16). We find that uncrowded sources in sample 1 show a homogeneous distribution over the whole sample volume, which can be clearly seen in Figure 9a (see also Figs. 15 and 16 presented below; note that many distant sources are missed due to flux limits). This suggests that some isolated regions of matter have been in isolation from the very beginning, and later...
on they will form galaxies by themselves. Crowded and very
crowded galaxies in sample 1 are much less homogeneously distrib-
uted (see Figs. 9b and 9c). This might be due to the continuous
action of gravity, which pulls these galaxies together, as expected
in many models. Governato et al. (1998) have already pointed
out that large concentrations are common in a universe domi-
nated by cold dark matter, and they are the progenitors of the rich
galaxy clusters seen today. It is interesting that filamentary fea-
tures are observed in a state of random distribution, formed
particularly by crowded sources (see Figs. 6b and 7b). This is
clearly shown in Figure 10. In addition, seeds of knots seem to be
present early in a random state as well (see Fig. 10). Agreeing
with the view of Governato et al. (1998), we suspect that it might
be these features that grow to the obvious clustering structures
observed at later times, where many presumed galaxies that were
originally uncrowded have joined. It seems that, in this process,
very dense regions move toward each other, and some of them
form large-scale structures themselves at later times. This is just
what is expected in the concentration process of dark matter (see
Fig. 1 of Governato et al. 1998). Revealed by recent observa-
tions, filamentary large-scale structures do exist at redshifts as
large as $z \sim 6$ (Ouchi et al. 2005).

Some voids are observed in the two-dimensional map of very
crowded sources in sample 1 (see Fig. 9c). These voids are also
detected in the spatial distribution of crowded sources in the sam-
ple, but they are less obvious (see Fig. 9b). On the contrary, we
find no obvious voids in the spatial distribution of uncrowded
galaxies in the same sample. Meanwhile, some embryonic voids
are seen in the two-dimensional plot of crowded sources in sam-
ple 3 (see Fig. 10a). This suggests that voids are likely volumes
within which no or very few very crowded sources are present,
and they are likely formed in embryo by fluctuations in the very
early epochs of the universe. It might be the continuous influence
of gravitation that pulls more crowded galaxies closer over the
course of time and at the same time leaves behind adult voids.
Shown in Figure 11 are the finer resolution maps showing the
details of the spatial distributions of sources of different crowd-
edness in sample 1 in two local regions of the volume concerned
(see Fig. 6a). Indeed, from the figure we find that if voids are
identified on the basis of the spatial distribution of very crowded
sources, then they are seen to be filled mainly with uncrowded
sources (for more detail, see Fig. 21).

The spatial density distributions of sample 1 calculated with
the first and second approaches are almost the same (see Figs. 6a
and 7a). Meanwhile, one might observe that the distributions of
$\kappa_i(r_{ccs})$ for sample 1 calculated with the two approaches are not
distinguishable (see Figs. 5a and 5c). This suggests that the
adopted values of $\rho_0(z)$ estimated directly from sample 2 and
evaluated by the empirical polynomial function do not provide a
significant difference in the analysis of the density distribution.
As discussed above, in examining the homogeneity properties of
the background samples, where the values of the correlation
dimension $D_2$ are checked, we arrive at the same conclusion.

---

**Fig. 8.**—Spatial (two-dimensional) distributions of sources of different crowd-
edness in samples (a) 1 and (b) 2, measured in the criterion clustering scale $r_{ccs}$.
The mean density $\rho_0 = 0.00247 (h^{-1} \text{Mpc})^{-3}$ evaluated from sample 1, which
is a constant with respect to redshift, is adopted to determine $r_{ccs}$ and to calculate $\kappa_i$. The meanings and the shades of gray of the symbols are the same as in Fig. 6.
For the definition of the coordinates, see Fig. 6. [See the electronic edition of the Journal for a color version of this figure.]

**Fig. 9.**—Spatial (two-dimensional) distributions of (a) uncrowded, (b) crowded,
and (c) very crowded sources in sample 1, measured in the criterion clustering scale $r_{ccs}$. This plot is a copy of Fig. 6a, where sources of different crowdedness
are presented in different panels. [See the electronic edition of the Journal for a color version of this figure.]
In the following, we discuss only the case of adopting the \( \rho_0(z) \) directly estimated from sample 2 (the first approach), where radial selection effects (known and unknown) are accounted for.

5. CLUSTERING PROBABILITIES

As long as the mean density \( \rho_0(z) \) is available, we are able to apply the FoF algorithm to identify clustering sources with the scale of \( r_{ccs}(z) \). To calculate the probability of creating a cluster with a given number of galaxies (in terms of statistics, the total number of members of a sample is always referred to as the size of the sample) and with a certain number density by chance, we perform a number of simulations, and in each run we sort out coherent clusters and calculate their neighborhood functions. Let the number of simulations showing clusters with at least \( n' \) galaxies and neighborhood functions larger than \( \kappa' \) be \( \Delta N \), and the total number of trials be \( N \). Then the probability of creating a cluster with the number of galaxies larger than \( n' \) and a neighborhood function larger than \( \kappa' \) by chance is estimated by dividing \( \Delta N \) by \( N \): 
\[
P\{n > n', \kappa > \kappa'\} = \Delta N/N.
\]

The details of our simulation analysis are summarized below (some of them can be found in previous sections): (1) We randomly select 59,497 redshifts from sample 2. These form a simulation sample which has the same number of sources as sample 1. In doing so, 226,302 redshifts corresponding to 226,302 sources contained in sample 2 have the same chance for being selected; in this way the redshift distribution of the simulation sample is the same as that of sample 2 (see Fig. 3b). (2) For each redshift that is selected, we modify it in the following way: Redshifts available in sample 2 are presented in the form
\[
z = n_0 + 0.1n_1 + 0.01n_2 + 0.001n_3 + 0.0001n_4,
\]
where \( n_0, n_1, n_2, n_3, \) and \( n_4 \) are numbers ranging from 0 to 9. The real redshift of a source might be different from the redshift so provided. For the sake of simplicity, we assume that \( 0.0001n_4 \) comes from \( 0.0001n_4 + x \), where \( x \) is a value satisfying \(-0.00005 < x < 0.00005 \) and ranging uniformly within this range. For instance, when selecting a redshift of 0.2981 from sample 2, we assume that this observed redshift arises from a real redshift ranging within \( 0.29805 < z < 0.29815 \), and then we modify the original redshift 0.2981 by adding a random value \( x \) to it (as assigned above, \( x \) is uniformly distributed within the range of \(-0.00005 < x < 0.00005 \)). We then get a new redshift set for the simulation sample. (3) Corresponding to each redshift of the simulation sample, we randomly select a sky position confined in the area of sample 1 and assign the redshift and the position to a source. In this way, we get 59,497 sources for the simulation sample. In doing so, the universe is assumed to be isotropic, and then the same area in the sky within the sample region (the area of sample 1; see Fig. 1) has the same probability of being selected. (4) We calculate \( r_{ccs} \) for each source in the simulation sample by applying the relation \( r_{ccs} = [2\rho_0(z)]^{-1/3} \), where the adopted \( \rho_0(z) \) is that directly measured from sample 2. (5) For this
simulation sample we apply the FoF algorithm with the \( r_{ccs} \) obtained above to sort out clusters. For each pair of sources we always have two values of \( r_{ccs} \). The two sources are considered to be within the same cluster when the distance between them is smaller than the smaller of the two \( r_{ccs} \) values. (6) For each cluster we calculate \( \kappa(r_{ccs}) \) using the estimator in equation (7), where the sum is taken over the whole number of sources in the cluster concerned. (7) We perform a number of simulations and then, according to these simulations, estimate the probability for forming a cluster with a certain number of galaxies and a certain value of \( \kappa(r_{ccs}) \) by chance.

The most important reason for choosing \( x \) (\(-0.00005 \leq x < 0.00005\)) to add to the redshift provided in sample 2 is that this makes the redshift distribution of the selected sample much less scattered than that of sample 2 (see Fig. 3b). In this way, the noise caused by the unmeasurable digital number of the redshifts is reduced. In doing so, we do not use the real redshift errors, since they are not taken into account when we deal with sample 1. In fact, as shown by Figures 6b and 7b, the spatial distributions of samples 3 and 4 are not distinguishable, although their redshift distributions are quite different in finer redshift intervals (see Fig. 3). This is also seen in the \( \kappa(r_{ccs}) \) distribution of sample 1, where the difference of redshifts (which is much more obvious than the difference between the real redshift errors and the range of \( x \) discussed above) does not lead to significantly different results (see Figs. 5a and 5c). In addition, one can check that replacing \(-0.00005 \leq x < 0.00005\) with the real redshift errors in the simulation analysis does not lead to an observable difference.

Displayed in Figure 12 are the probability contours of coherent clusters formed by chance (here \( N = 5000 \)). It shows that it is more difficult for clusters with larger numbers of galaxies and larger number densities (larger \( \kappa \)) to be formed. The probability of forming coherent clusters with more than 350 galaxies is less than that of \( 3 \sigma \).

Our analysis shows that if a cluster with number of galaxies \( n \) larger than 10 and \( \kappa \) larger than 5.41 could be formed by chance, the probability must be less than that of \( 3 \sigma \), \( P(3 \sigma) \). Under the same requirement of probability, other typical pairs of \( (n, \kappa) \) are \((100, 2.24)\), \((200, 1.10)\), and \((300, -0.99)\). This shows that for clusters with a very large number of galaxies (say, \( n > 300 \)), the value of \( \kappa \) is small, and it could even be negative. This suggests that it is very difficult for clusters with a very large number of galaxies to be formed by chance. Even when their mean densities are smaller than the average (i.e., \( \kappa < 0 \)), it is still very difficult for them to be formed. Thus, when a cluster containing a number of galaxies larger than 300 is identified in an area as large as that of sample 1, it is unlikely that this cluster is formed by chance.

6. IDENTIFYING AND CLASSIFYING COHERENT CLUSTERS

Applying our method to samples 1 and 3, we get entirely different sets of coherent clusters. In sample 1 we detect 67 coherent clusters under the condition that if any of them is formed by chance then the probability will be less than that of \( 3 \sigma \). Listed in Table 2 are the parameters of these coherent clusters. Here the scale of a coherent cluster is defined as the largest value of the distance measured for each pair of sources in the cluster. The smallest neighborhood function is \( \kappa = 1.79 \). For the other 66 clusters, \( \kappa > 2 \). The largest number of galaxies is 12,966, and the largest scale is 357 \( h^{-1} \) Mpc, which belong to different coherent clusters. For all these clusters \( \kappa > 1 \), suggesting that the number density for any of these clusters, measured for each source within the scale of \( r_{ccs}(\sigma) \), is larger than twice the mean of the background sample. For sample 3 the largest number of galaxies in a detected coherent cluster is 135, and the neighborhood function of this cluster is \( \kappa = 1.22 \). According to Figure 12, if the number of galaxies of a cluster is larger than 134, then when the neighborhood function of this cluster is larger than 1.29, the probability of forming this cluster by chance is less than \( P(3 \sigma) \). For the cluster concerned, while the first condition is satisfied, the second is not. Thus, the probability of forming this cluster by chance is larger than that of \( 3 \sigma \). The scattering distribution in the \( n-\kappa \) plane of the coherent clusters detected from sample 3 is shown in Figure 13. As seen in the figure, there are no coherent clusters detected within the 3 \( \sigma \) probability region for sample 3. Also displayed in Figure 13 is the scattering distribution of the coherent clusters detected from sample 1. One finds from this figure that almost all coherent clusters detected from sample 3 are distributed beyond the 1 \( \sigma \) probability region in the \( n-\kappa \) plane. While a sufficient number (67) of coherent clusters sorted out from sample 1 are situated within the 3 \( \sigma \) probability region, there are some other coherent clusters detected from the sample existing within the 1 and 3 \( \sigma \) probability contours.

We define three kinds of coherent clusters: prominent clusters, \( P < P(3 \sigma) \); less prominent clusters, \( P(3 \sigma) \leq P < P(1 \sigma) \); and weak clusters, \( P(1 \sigma) \leq P \). Spatial (two-dimensional) distributions of the three kinds of clusters, as well as the unclustered sources for the two samples, are shown in Figures 14 and 15. We find that the unclustered sources of both the observed and background samples are homogeneously distributed over the whole space concerned (see also Figs. 16a and 16b). They appear like low relative density galaxies (\(-1 \leq \kappa_i < 1 \)); see Fig. 9a), and indeed, almost all of them are in fact low relative density galaxies (see Figs. 15a and 15b). As suggested in Figures 15a and 15b, the number of unclustered sources detected from sample 3 is larger than that from sample 1. This is expected since many original unclustered regions (the presumed galaxies) might be pulled toward a nearby relatively dense region (a presumed cluster) over the course of time due to the continuous action of gravity. Figure 15 shows that the majority of medium and high relative density sources (those of \( 1 \leq \kappa_i < 3 \) or \( \kappa_i \geq 3 \)) are members of clusters (weak, less prominent, or prominent clusters). Compared with the corresponding panels in Figure 16 we find that sources of weak clusters in the background sample are homogeneously distributed over the whole volume of the sample as well, and the spatial distribution of sources of this kind in the observational sample is quasi-homogeneous (see panels c and d in Figs. 15 and 16).
As expected, as shown in Figures 15 and 16, the redshift distributions of the unclustered, weakly clustered, less prominently clustered, and prominently clustered sources of samples 1 and 3 are displayed in Figure 16. As expected, the redshift distributions of the unclustered, weakly clustered, less prominently clustered, and prominently clustered sources of the observed sample is consistent with that of the whole sample, which is due to the very small number of this kind of source, i.e., 0.568% of the sample. While the redshift distribution of the unclustered sources of the observed sample is consistent with that of the background sample (see Fig. 16a), there is a slight deviation of the weakly clustered sources of sample 1 from that of the background sample, and an obvious deviation of the prominently clustered sources of sample 1 from that of the latter sample is observed. For the observed sample, the number of sources of less prominent clusters is much smaller than both the weak and prominent clusters. In addition, we find that for the observed sample there is a change from weak clusters to prominent clusters; the...
number of sources in between (sources of less prominent clusters) is relatively small.

According to this analysis, we suspect that there might exist three kinds of galaxies or galaxy groups in terms of clustering. One class is made up of isolated galaxies which are homogeneously distributed over the universe and are never included in any clusters. The positions of these galaxies would be at rest in the frame of the universe. Another class includes trivial coherent clusters which act like the isolated galaxies except that their number densities evolve with time. The third class contains prominent coherent clusters whose densities and numbers of galaxies grow continuously from very early epochs to the present. After the clustering process of all the galaxies in the universe ceases, the redshift distribution of the sources in these clusters would reflect a homogeneous spatial distribution when measured in a volume much larger than that associated with their typical scales. Measuring within a volume comparable to that associated with their typical scales, it would not be surprising if the redshift distribution showed inhomogeneous spatial distribution. Before the clustering process stops, some potential members of these clusters would look like trivial coherent clusters in terms of density (described by $\kappa$) and number of galaxies, and this would make the redshift distribution of apparent trivial coherent clusters (weak clusters defined in this paper; see Fig. 15c) deviate from a homogeneous spatial distribution (see Fig. 16c). In this scheme, the less prominent clusters defined in this paper would be potential members of the third class, and they would change their identities with time. According to this interpretation, once a prominent cluster is formed it will maintain its identity. This explains why there is a turnover change from the number of weak clustering sources to that of prominent clusters observed above.

As revealed in Figure 13, there are two approaches for a weak cluster to become a prominent cluster: either by increasing its density (described by $\kappa$) or by increasing the number of members. In the first approach the members of the cluster get closer, while in the other approach the cluster gets coherently connected with other clusters. For an isolated cluster, when its relative density stops growing, its identity will not change any more.

The reasons that we propose the above scheme to interpret the clustering process of galaxies are that (1) we find it natural to explain what we observe in the above clustering plots; (2) it looks simple; (3) it can provide unambiguous predictions in the clustering process which can be checked later; (4) we find no simpler schemes accounting for our results.

In the above analysis we first classified four kinds of sources according to their clustering properties available from observation: unclustered sources (which do not appear in Fig. 13), weakly clustered sources (which are below the $1/\sqrt{C_2}$ contour in Fig. 13), less prominently clustered sources (which are within the area confined by the $1$ and $3/\sqrt{C_2}$ contours in Fig. 13), and prominently clustered sources (which are above the $3/\sqrt{C_2}$ contour in Fig. 13). Later we proposed a scheme with three kinds of galaxies or galaxy groups (isolated galaxies, trivial coherent clusters, and prominent coherent clusters) to explain what we observe in the statistical analysis. The four kinds of sources discussed above might change

Fig. 13.—Plot of $\kappa$ vs. number of galaxy members of coherent clusters identified in samples 1 (crosses) and 3 (plus signs), where the probability contours in Fig. 12 are also presented.

Fig. 14.—Spatial (two-dimensional) distributions of clustered sources of samples (a) 1 and (b) 3. The dark-gray color stands for unclustered sources, the medium-light-gray color represents sources of weak clusters, the light-gray color denotes sources of less prominent clusters, and the medium-dark-gray color symbolizes galaxies of prominent clusters. For the definition of the coordinates, see Fig. 6. [See the electronic edition of the Journal for a color version of this figure.]
their identities during the clustering process, and when the clustering process ceases, each of them will belong to one of the three kinds of galaxies or galaxy groups (i.e., isolated galaxies, trivial coherent clusters, or prominent coherent clusters).

7. STRUCTURE OF LARGE PROMINENT COHERENT CLUSTERS

Here we show the structure of the two largest prominent coherent clusters identified in sample 1. According to Table 2, the coherent cluster with the largest number of galaxies is cluster 1, which contains 12,966 galaxies and extends to $281 h^{-1}$ Mpc; the one with the second largest number of galaxies is cluster 2, which contains less galaxy members (7788) but has the largest scale, $357 h^{-1}$ Mpc. The scale of cluster 1 is the distance between the two sources TGN 141Z157 and TGN 353Z218, and that of cluster 2 is the distance between TGN 163Z153 and TGN 276Z102.

Displayed in Figure 17 are the two-dimensional positions (relative to that of other galaxies of sample 1) of the two coherent clusters. Shown in Figures 18 and 19 are their two- and three-dimensional structures, viewed from various angles in the latter case.

As shown in Figure 17, the two clusters are quite close. The scales of the two coherent clusters are comparable with the whole volume of sample 1 (the upper limit in redshift in sample 1 is
sources are found? Figure 21 shows the structures of the two clusters together with that of other prominent clusters and the spatial distribution of galaxies of different crowdedness in sample 1 within a smaller area enclosing the structures. We find from Figure 21a that, for the majority of voids detected in the structures of the two clusters, there are indeed no or very few crowded sources present. This can be clearly seen in Figures 21d and 21e. Within the apparent void at \((x, y) \sim (-300 \, h^{-1} \text{ Mpc}, -50 \, h^{-1} \text{ Mpc})\) in Figure 21a, there are some very crowded sources. We suspect that this might not be a real void but instead might be due to a projection effect. In other words, the inner structures observed within the apparent void might not be close to the two clusters in three-dimensional space, but they appear so in the projected two-dimensional space. Indeed, as revealed in Figures 21b and 21c, inside the apparent void some structures of prominent clusters are present. Shown in Figure 21d, there are indeed few very crowded sources outside the structures of prominent clusters, and it is these galaxies that form the core and build the frame of the structures. Crowded sources tend to be gathered around the structures, as seen in Figure 21e. Just as pointed out above, uncrowded sources are seen to be homogeneously distributed over the structure area, and they are indeed detected within the voids, as well as within the frame of the structures (see Fig. 21f). In a recent investigation of voids, Patiri et al. (2006) showed that faint galaxies populating the voids are clustered in small groups and filaments. This is observed in Figure 21f. In addition, we find that the apparent filaments of uncrowded galaxies do not stretch along the frame of the structure of the prominent clusters. On the other hand, filaments built of very crowded and crowded sources do stretch along the frame of the structure, as is observed in Figures 21d and 21e. As filamentary features could appear in a state of random distribution (see the discussion in §4), we suspect that the apparent filaments of low-density galaxies might mainly be caused by fluctuations.

8. THE LARGEST STRUCTURE DETECTED IN A LARGER DATA SET OF THE 2dF SURVEY

One might observe that, in the method adopted above, while the calculation of the relevant probability and the investigation of number density distribution depend on the exact volume of the adopted sample, the sorting method (i.e., the FoF algorithm) is independent of the volume. This enables us to identify the largest scale structure of a sample of the same survey for which the borders are not well defined and the volume is not available. Although the probability of forming such a structure by chance is difficult to evaluate, at least the following questions might be addressed. How large is the scale of the largest structure identified from the whole 2dF survey by the FoF algorithm? Is it that some unconnected structures identified in a smaller volume can be coherently connected in a larger volume, which is highly expected? Is the coral-type structure detected in the space of sample 1 merely a local phenomenon? (Or could it be found in other areas of the survey?)

Here, we apply the sorting method proposed above to the whole 2dFGRS data set. To match the basic criteria of sample 1, we simply adopt sample 2 for our analysis. The \(\rho_0(z)\) used to determine the \(r_{cc}\) is that directly measured from sample 2 (i.e., \(\rho_1\) in Table 1).

The largest number of galaxies in the coherent clusters identified in sample 2 is 41,813, and the scale of this cluster is \(659 \, h^{-1} \text{ Mpc}\), which is the largest one detected with the adopted method. The second largest cluster contains 32,188 galaxies, and its scale is \(474 \, h^{-1} \text{ Mpc}\). The two-dimensional positions (relative to the other galaxies of sample 2) of the two coherent clusters are illustrated in Figure 22, which shows that the structures of the
two groups of galaxies are situated in two distinct areas of the sky. Comparing it with Figure 17 we find that the two largest coherent clusters identified from sample 1, clusters 1 and 2, now become a single one, the second largest cluster identified from sample 2 (the one shown on the left of Fig. 22). Besides the sources in clusters 1 and 2, some other sources are also included in the second largest cluster. This is expected, since the area of sample 1 is well inside that of sample 2 in that direction of the sky (see Fig. 1), and some bridge sources outside the area of sample 1 might exist and then would connect clusters 1 and 2 together, as well as connect other sources or clusters to them.

The right side of Figure 22 shows that large-scale structures could extend to the volume limit of the adopted sample (the largest redshift detected in the right-side structure, the largest coherent cluster detected from sample 2, is \( z = 0.2646 \)). (For the left-side structure, the largest redshift detected is \( z = 0.1969 \).) This indicates that large-scale structure is not a proprietary feature of nearby galaxies. Instead, it could have formed much earlier, when the space density of galaxies was sparse (see Fig. 8a). In fact, it has been found in recent observations that large-scale structures are common at redshifts as large as \( z \sim 3 \) (Steidel et al. 1998). Large-scale structures made up of protoclusters could have formed even at epochs as early as \( z \sim 6 \) (Ouchi et al. 2005).

The structure of the largest cluster detected from sample 2, that on the right of Figure 22, is displayed in Figures 23 and 24, where more details are visible (also shown in Fig. 23 is the

---

**Fig. 19.**—Three-dimensional map of the structures of the two largest coherent clusters, where the light-gray color represents cluster 1 and the dark-gray color denotes cluster 2. Here, a relatively lower resolution map is presented so that the solid structure is noticeable. For the definition of the coordinates, see Fig. 6. [See the electronic edition of the Journal for a color version of this figure.]
two-dimensional structure of the second largest cluster). Figure 23, which is a high-resolution map, shows that the topology of the structure is coral-like as well.

Although the probability for forming the two largest coherent clusters by chance is not available due to the ill-defined borders of sample 2, we tend to believe that the two clusters are not formed by chance. According to Figure 12, large numbers of galaxies of coherent clusters are scarcely formed by chance, even though their mean relative densities are small. In sample 3 the largest number of galaxies is much smaller than 300 (see Fig. 13), but the largest number detected from sample 1 is 12,966, which is 43 times 300. The largest and the second largest numbers of galaxies in the coherent clusters identified from sample 2 are 139 and 107 times 300, respectively. It is unlikely that the enhancement of the volume from sample 1 to sample 2, where the number of galaxies of the latter is about 4 times the former, would shift the 3σ probability contour in Figure 13 to a number of galaxies as large as that of the second largest cluster, 32,188.

9. DISCUSSION AND CONCLUSIONS

Our analysis is performed in redshift space, where the peculiar velocity of individual sources is ignored. As is known, in
measuring the distance of an object, a major problem is to distinguish the dynamical velocity from the observed redshift (see, e.g., Gellar & Huchra 1989; Guzzo 2002). Generally, velocity distortion shrinks overdense regions and inflates underdense ones (Kaiser 1987; Martínez & Saar 2002a). As shown in Figure 13, correction for this effect would turn some weak clusters of sample 1 into less prominent clusters, since their number densities (represented by $\kappa$) would be larger. In contrast, the correction would turn some prominent clusters into less prominent ones, since their number densities would be smaller.

When the redshift distortion is removed, the pattern of clustering observed above might be altered. However, as revealed by Figure 18, the possible change in the two largest structures of sample 1 would not be so severe, since the structures themselves are firmly built on a very dense framework.

The measurement uncertainty in the redshift for the 2dFGRS is 85 km s$^{-1}$ (Colless et al. 2001), which corresponds to $\Delta z = 2.8 \times 10^{-4}$. The line-of-sight positions are certainly affected by errors in redshift. Does this play an important role in identifying clusters when the FoF algorithm is applied? The answer is yes when the redshift concerned is $<0.001$. Accordingly, as revealed by Table 2, clusters 14, 16, 22, 33, 47, 52, 54, 61, and 64 are certainly affected by the redshift measurement uncertainty. When a much smaller redshift uncertainty is available, some of these clusters might be disassembled, while some of them might get merged into other clusters. However, as illustrated in Figure 3, most sources in sample 1 have redshifts much larger than the uncertainty, and thus, the effect of redshift error must be very limited in identifying clusters among them. In particular, the largest clusters discussed above are located in relatively high redshift regions (see Fig. 22 and Table 2). They are much less affected by this uncertainty.

Colless et al. (2001) pointed out that the survey magnitude limit of 2dFGRS varies slightly with position on the sky. In their SGP strip the median limiting magnitude is $b_J = 19.40$, with an rms of 0.05 mag; in their NGP strip the median limiting magnitude is $b_J = 19.35$, with an rms of 0.11 mag. To show the effect of this, we repeat the grouping analysis performed above by adopting a brighter magnitude cut so that the sample selected will be more complete in magnitude. According to Colless et al. (2001), we take the limiting magnitude as $b_J = 19.24$ and consider only galaxies brighter than this. Under this condition, there are 198,647 sources left in sample 2 (87.8%) and 55,846 galaxies remaining in sample 1 (93.9%). Due to the cut in the total number of samples, the straightforward prediction is that all the clusters identified above will have a smaller number of members and a smaller value of $\kappa$, and this is found to be true. Cluster 1 now has only 12,391 sources with $\kappa = 5.90$. For cluster 2, the number of members now becomes 6237, and the value of $\kappa$ is 3.97. The scale of cluster 1 remains unchanged, while that of cluster 2 now becomes 303 h$^{-1}$ Mpc. This is not surprising, since faint sources are always at large distances. Note that cluster 2 is identified at a larger distance than cluster 1. The spatial distributions of the two clusters are shown in Figure 25, where for the sake of comparison the original members of the two clusters are also presented. One finds that while cluster 1 remains almost the same, cluster 2 is obviously affected. Some original substructures of cluster 2 are dismissed, while some new substructures are attached. We arrive at the same conclusion as obtained earlier: i.e., clusters can stretch to large distances. (What would happen if we continuously adopted brighter magnitude limits? Currently we have no answers to this. It deserves an intensive investigation.) Besides the number and density of clusters, the new magnitude limit might also affect the probability contours of sample 1. Since the total number of the sample becomes smaller than the previous one, clusters identified with the same algorithm would have smaller numbers and would be less dense. In this way, the contours shown in Figures 12 and 13 would shift leftward. As revealed by Figure 13, clusters 1 and 2 are far away from the $3 \sigma$ contour, even though they have smaller numbers than before due to the brighter limit. The conclusion that the probability of picking the two clusters by chance is much less than that of $3 \sigma$ is maintained.

Some authors have proposed that the three dimensional topology of large-scale structures is sponglike (Gott et al. 1986; Vogelezang et al. 1994; Colley et al. 2000). Indeed, when plotting the structures of the two largest clusters identified from sample 1 in a lower resolution map, this type of feature emerges (see Fig. 26). It suggests that the large-scale structure of the local universe is intrinsically coral-like, but it could also be sponglike if the resolution of the map were low enough.

Gott et al. (2005) have already pointed out that some larger structures are found if larger samples are considered. This is natural and is indeed observed in the analysis of sample 2, where the two large-scale structures detected in a smaller sample (sample 1) merge and a much larger structure is identified (see Fig. 22). (Since the volume of sample 2 occupies only a small fraction of the nearby space, the above argument suggests that we might be able to observe a structure with a scale as large as $\sim 1000$ h$^{-1}$ Mpc in the local universe as long as the whole sky is surveyed.)

As illustrated in Figure 21, some of the coherent clusters and unclustered sources appear to connect with each other in the two-dimensional view. This is indeed true. When one repeats the above analysis of sample 1 under the condition that the contribution of
the vertical length (i.e., the contribution of the coordinate \(z\) defined in Fig. 6) is ignored, one identifies a much larger coherent cluster than the largest one obtained above (the details of the analysis are omitted). This suggests that some large-scale structures observed in a two-dimensional map might not be coherent when the vertical length is taken into account. As shown in Table 2, the mean redshift of cluster 1 is 0.0864. At this redshift, the vertical length confined by the adopted window of sample 1 [say, \(-4.8^\circ < \text{decl.}(\text{J2000}) < 1.0^\circ\)] is 25.7 \(h^{-1}\) Mpc, while the adopted criterion clustering scale at this redshift is \(r_{\text{ccs}} = 3.07 \, h^{-1}\) Mpc. The former is about 8 times the latter. Keeping this in mind, it is expected that some structures which are seen connected in the two-dimensional map are in fact separated.

Note that \(r_{\text{ccs}}\) might be different for different surveys. It is expected that the number of galaxies we will eventually be able to see will be larger by a factor of 2.36 than the number that is observable today (Gott et al. 2005). If sample 1 is enlarged by 2 times, \(r_{\text{ccs}}\) would probably shrink to a smaller scale. This does not necessarily mean that this would lead to a smaller scale of the largest structure identified with \(r_{\text{ccs}}\). One reason is that the enlarging of the sample makes the sources more crowded, and then it is easier for separated sources to be connected. Another reason is that the largest structure is constructed with the framework made up of very crowded sources (with several times the number density of the background data). The enlargement of the sample (only 2 times larger) would not be able to change this situation.

It should be pointed out that structure scales as large as 357 \(h^{-1}\) Mpc obtained from sample 1 and as large as 659 \(h^{-1}\) Mpc obtained from sample 2 depend on the criterion clustering scale, \(r_{\text{ccs}} = [2\rho_0(z)]^{-1/3}\), adopted in this paper. It is obvious that when a smaller \(r_{\text{ccs}}\) is adopted, a smaller scale for the largest structure identified from the samples will be found, and when a larger one is taken, one will get a larger scale for the largest structure. Thus, while talking about the largest structure scale one should refer to...

---

**Fig. 24**.—Three-dimensional map of the structure of the largest coherent cluster identified from sample 2. Here, a relatively lower resolution map is presented so that the solid structure is noticeable. For the definition of the coordinates, see Fig. 6. [See the electronic edition of the Journal for a color version of this figure.]
the corresponding criterion clustering scale. Due to this and the fact that the criterion clustering scale adopted in this paper varies with redshift, a comparison between the largest structure scale in our analysis and that in other previous studies is not meaningful. It is desirable to construct a standard criterion clustering scale and then compare the largest structure scale obtained from various samples or from different regions of the universe. Although we are unable to compare our results with others, the algorithm adopted in this paper tends to pick out larger structures than what have been identified previously when similar criterion clustering scales are adopted. It is because we identify in this paper large-scale structures in the comoving frame of the universe rather than in the observer frame. It is possible that some galaxies excluded in the previously identified large-scale structures due to their large redshifts have become members of some large-scale structures identified in this paper when they are located in a relatively dense region (see Fig. 22).

Martinez et al. (1998) searched for the scale of homogeneity by applying the $K$-function to galaxy catalogs. They detected the fingerprint of the transition to homogeneity in all cases considered. Amendola & Palladino (1999) analyzed volume-limited sub samples of a redshift survey to search for the scale of homogeneity and found that the survey shows a trend of homogeneity at large scales. As shown in Figure 4, one finds a trend toward homogeneity as well. We argue that at least what is revealed by Figure 4 cannot be interpreted as the existence of a transition to homogeneity at a large scale, since that scale is comparable to the maximum scale of the adopted sample (see the caption of Fig. 4). As pointed out in § 3, we suspect that the flux limit, as well as a possible evolutionary effect, might be the main factors that block the detection of the real scale of homogeneity. Neither previous methods nor the neighborhood function introduced in this paper can solve this problem. Only when this effect is removed will one be able to determine the homogeneity scale with galaxy redshift surveys.

As suggested above, filamentary features are observed in a state of random distribution, formed by crowded sources (see Fig. 10). This is in agreement with the previous prediction that galaxies preferentially form in large-scale filamentary or sheet-like overdensities in the early universe, which have been detected at distances as far as $z = 3.1$ (Matsuda et al. 2005). When sorting out all sources of the background sample with $r_{\text{oss}}$, we get many weak clusters and a few less prominent clusters (see Figs. 15d and 15f). According to the definition, these clusters are likely the so-called protoclusters (Venemans 2005). The latter were identified at $z \sim 4$ in recent observations (Venemans et al. 2002; Miley et al. 2004; Intema et al. 2006).

According to the above analysis we conclude that (1) the probability of forming the large-scale structures detected from sample 1 with our method by chance is very small; (2) the phenomenon of clustering is dominant in the local universe; (3) coherent clusters with scales as large as $357 h^{-1}$ Mpc and numbers of galaxies as large as 12,966 are identified in sample 1 well within the 3-$\sigma$ confidence level; (4) there exist some galaxies which are not affected by the gravitation of clusters, and therefore are likely to be at rest in the comoving frame of the universe; (5) filamentary features could appear in a state of random distribution; (6) voids are likely volumes within which very crowded sources are not present, and they are likely formed in embryo by fluctuations in the very early epochs of the universe; it might be the continuous action of gravitation over the course of time that pulls crowded galaxies closer and at the same time leaves behind adult voids; (7) large-scale structures are coral-like, and they are likely made up of smaller ones; in forming the former, the latter are likely connected by their filaments; (8) very crowded sources are mainly distributed within the structure of prominent clusters, and it is they that form the framework of a large-scale structure.
