Mon. Not. R. Astron. Soc. **335,** 977–983 (2002)

Scaling of voids and fractality in the galaxy distribution

José Gaite and Susanna C. Manrubia★

Centro de Astrobiología, CSIC-INTA, Ctra de Ajalvir Km 4, 28850 Torrejón de Ardoz, Madrid, Spain

Accepted 2002 May 13. Received 2002 May 10; in original form 2002 March 20

ABSTRACT

We study here, from first principles, what properties of voids are to be expected in a fractal point distribution and how the void distribution is related to its morphology. We show this relation in various examples and apply our results to the distribution of galaxies. If the distribution of galaxies forms a fractal set, then this property results in a number of scaling laws to be fulfilled by voids. Consider a fractal set of dimension D and its set of voids. If voids are ordered according to decreasing sizes (largest void has rank $R = 1$, second largest $R = 2$ and so on), then a relation between size Λ and rank of the form $\Lambda(R) \propto R^{-z}$ must hold, with $z = d/D$, and where d is the Euclidean dimension of the space where the fractal is embedded. The physical restriction $D < d$ means that $z > 1$ in a fractal set. The average size $\bar{\Lambda}$ of voids depends on the upper (Λ_u) and the lower (Λ_l) cut-off as $\bar{\Lambda} \propto \Lambda_u^{1-D/d} \Lambda_l^{D/d}$. Current analyses of void sizes in the galaxy distribution do not show evidence of a fractal distribution, but are insufficient to rule it out. We identify possible shortcomings of current void searching algorithms, such as changes of shape in voids at different scales or merging of voids, and propose modifications useful to test fractality in the galaxy distribution.

Key words: methods: statistical – galaxies: general – large-scale structure of Universe.

1 INTRODUCTION

The morphological properties of the distribution of galaxies are commonly analysed by means of the correlation functions of this distribution – chiefly, the two-point correlation function (Peebles 1980). This correlation function is well fitted by a power law up to some scale (Peebles 1980). Within this range, the coarse-grained galaxy density exhibits large fluctuations associated with various structures – namely, galaxy clusters and superclusters of diverse forms, and voids. Most studies of galaxy structure have focused on clusters and superclusters, but the presence of large voids was noted long ago and the size of the largest voids detected has steadily grown (Einasto, Jõeveer & Saar 1980). The analysis of voids is a subject of current interest in cosmology (Einasto, Einasto & Gramann 1989; Vogeley, Geller & Huchra 1991; El-Ad, Piran & da Costa 1997; Müller et al. 2000; Hoyle & Vogeley 2002).

On the one hand, the analysis of correlation functions and the hierarchical structure of clusters and superclusters provides evidence for a self-similar *fractal* structure (Coleman & Pietronero 1992; Sylos Labini, Montuori & Pietronero 1998), although the scale of transition to a homogeneous universe is still a matter of debate (Guzzo 1997; Wu, Lahav & Rees 1999; Chown 1999; Martínez 1999). On the other hand, the scaling properties of voids are much less studied, but scaling of certain quantities has been put forward as an indication of self-similarity (Einasto et al. 1989).

Here, we begin by studying the void properties of fractal distributions in general. Self-similarity is the most obvious property and is related to the *fractal dimension* but there are other properties worth considering, such as *lacunarity* (Mandelbrot 1977), which we define in Section 2. We show in examples how to perform a void analysis to obtain the fractal dimension and other morphological properties. Then we proceed to compare with current void analyses of galaxy catalogues, pointing out their relation with our method and, according to this method, the conclusions that can be extracted from these catalogues.

Typically, voids are extracted from galaxy catalogues by using some *void detection algorithm*. These algorithms provide us with a list of voids, ranked by decreasing size. Therefore, these lists are suitable for rank-ordering techniques common in statistics (Zipf 1949; Sornette 2000). In particular, *Zipf's law*, which is a rank-ordering power law, is often indicative of fractal behaviour. In our case, a power-law cumulative distribution of voids is expected for a geometric fractal (Mandelbrot 1977). This cumulative distribution corresponds to a rank-ordering Zipf law (with different exponent).

We shall begin studying the rank ordering of voids in terms of simple geometric fractals (namely, Cantor sets – for which the Zipf law can be easily proved). Furthermore, we will consider some suitable two-dimensional fractals, where the problem of definition of voids arises. We will extrapolate the results to three-dimensional fractals. In all cases, we will refer to *pure fractals*, which can be characterized by a single exponent (i.e. their fractal dimension). We will not discuss the more complex case of *multifractal sets*, where

★E-mail: cuevasms@inta.es

a whole spectrum of singularity exponents is required to accurately describe their geometrical properties (Falconer 1990). Finally, we apply the previous results to current galaxy void catalogues.

2 ZIPF'S LAW AND FRACTALITY IN CANTOR SETS

Consider a set of quantities $\{\Lambda_k\}$ corresponding to $k = 1, 2, \ldots$ measures of a phenomenon. It is common usage to measure the probability distribution $p(\Lambda)$, i.e. the probability of finding an event of size Λ, in order to quantify the statistical properties of the system. An alternative way to carry out a similar quantification is provided by the *rank-ordering technique*, introduced by Zipf in the fifties (Zipf 1949). This procedure highlights the properties of the large values of Λ: the largest value is assigned rank $R = 1$, the second largest has $R = 2$, and so on. The function $\Lambda(R)$ conveys information equivalent to $p(\Lambda)$. In particular, if $\Lambda(R) \propto R^{-z}$, then $p(\Lambda) \propto \Lambda^{-\alpha}$, with $\alpha = 1 + 1/z$. This relation can be explained as follows. Note that $p(\Lambda)$ is the fraction of voids with size Λ. Hence, the total number of voids with size larger than or equal to Λ [which corresponds to the function $R(\Lambda)$] is proportional to the accumulated distribution $p(\Lambda)$, i.e.

$$R(\Lambda) \propto \int_{\Lambda}^{\infty} p(\Lambda)\, \mathrm{d}\Lambda. \tag{1}$$

If $p(\Lambda) \propto \Lambda^{-\alpha}$, direct integration returns $R(\Lambda) \propto \Lambda^{1-\alpha}$. Inverting it, we obtain the reported relation between z and α.

In order to illustrate the relation between Zipf's law for void sizes and the geometrical properties of a matter distribution, we begin with Cantor-like sets defined in the unit interval. If we restrict to deterministic fractals, a number of relevant quantities can be exactly calculated and clearly put in correspondence with each other.

A deterministic Cantor set is generated by an iterative procedure. Its *generator* is characterized by three independent quantities. First, $r < 1$ is the scaling factor. Usually, the unit interval is divided into $1/r$ pieces of equal length. Of these, N intervals remain for the process to be repeated and $(1/r - N)$ are eliminated. These two quantities completely define the fractal dimension of the asymptotic set, namely $D = -\log N/\log r$ (Mandelbrot 1977). None the less, there are different ways in which the N intervals can be chosen (in particular, some of them could be adjacent). Therefore, there is still a degree of freedom which translates into a variable number $m < N$ of voids (or gaps) in the generator. The more adjacent intervals, the fewer gaps and smaller m.

The effect of the parameter m in the morphology of the fractal set is quantified through an appropriate measure of *lacunarity*, i.e. the quality of having large voids (for a given sample size). Fig. 1 shows three examples of generators and the first iteration for sets with $N = 5$, $r = 1/9$ (hence with the same fractal dimension) and different m. The classic triadic Cantor set has $N = 2$, $r = 1/3$, and $m = 1$. Finally, note that the average length $c(r, N, m) \geqslant 1$ (in units of r) of the m initial voids can be obtained from the relation

$$cm + N = \frac{1}{r}, \tag{2}$$

where the function $c(r, N, m)$ gives an estimation of the degree of 'adjacency' of voids of size r. In the following, we write only c for simplicity.

After the independent quantities N, r and m have been introduced, we can turn to the explicit calculation of Zipf's law and related quantities. In the first iteration of the deterministic process, we have m voids with ranks from $R_1 = 1$ to m and length $\Lambda_1 = cr$. In the second iteration, there will be mN voids occupying ranks from $R_2 = 1 + m$ to $m + mN$, and their typical length is $\Lambda_2 = cr^2$. In general, in the ith iteration there are mN^{i-1} voids of average size

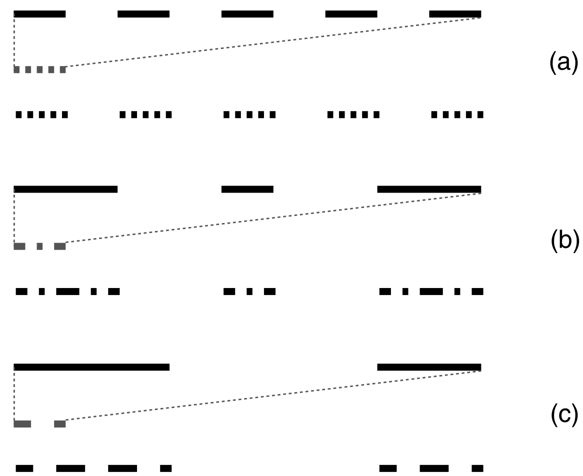

Figure 1. Three generators for Cantor-like fractals in the unit interval and the first iteration of the algorithm. We show the generator and the scaling rule producing the fractal set: N repetitions of the scaled set are used to construct the next iteration. The three examples shown have a variable number of gaps in the generator: (a) $m = 4$, (b) $m = 2$, (c) $m = 1$. Other parameters are shared: $N = 5$, $r = 1/9$. Hence, these three fractals have the same fractal dimension $D = \log 5/\log 9$ but different lacunarity.

$$\Lambda_i = cr^i \tag{3}$$

and ranking from

$$R_i = 1 + m\frac{N^{i-1} - 1}{N - 1} \tag{4}$$

to $R_i + mN^{i-1} - 1 \equiv R_{i+1} - 1$. One can verify that the rank of the first and the last void in each size class scales in the same way with the parameters of the system (the function $\Lambda(R)$ is step-shaped with steps of equal length in logarithmic scale – see the appendix). For the sake of clarity, we will use only the value R_i to calculate the explicit form of Zipf's law, defined in parametric form by equations (3) and (4). Eliminating the parameter i and arranging terms we get (for large R),

$$\Lambda(R) \approx f(r, N, m) R^{-1/D}, \tag{5}$$

as shown in the appendix, where D is the fractal dimension of the set and

$$f(r, N, m) = \frac{1 - rN}{m}\left(\frac{N - 1}{m}\right)^{-1/D}. \tag{6}$$

Mandelbrot (1977) introduced the gaps' length distribution N_r, defined as the cumulative number of gaps with length larger than a certain given scale Λ, and proposed that it is a power law with exponent $-D$ ($N_r \propto \Lambda^{-D}$). Now notice that the rank R_i is defined by adding the total number of gaps larger or equal than the ith gap. Hence, the gaps' length distribution corresponds to $R(\Lambda)$, which we can get through inversion of equation (5):

$$R(\Lambda) = F(r, N, m)\Lambda^{-D}, \tag{7}$$

and where the prefactor

$$F(r, N, m) = f(r, N, m)^D = \frac{m}{N - 1}\left(\frac{1 - rN}{m}\right)^D \tag{8}$$

is (a measure of) the lacunarity of the fractal set. Indeed, $F \propto m^{1-D}$ grows with m – so the smaller F is, the more 'lacunar' a fractal becomes. F ranges from $(1 - rN)^D/(N - 1) < 1$ for $m = 1$ to one for $m = N - 1$. We conclude that F^{-1} is a measure of lacunarity, in accord with Mandelbrot (1977).

However, there are other measures of lacunarity, and Mandelbrot actually concluded that it might be best to consider the fluctuations of the mass function $\mathcal{M}(\mathcal{R})$ (Mandelbrot 1977) (defined as the mean mass inside a ball of radius $\mathcal{R}$ centred on a point, $\mathcal{M}(\mathcal{R}) \propto \mathcal{R}^D$). This measure of lacunarity is related to the three-point correlation function and has been the one most employed (see Blumenfeld & Ball 1993, and references therein).

A particular case of the relations above is provided by fractals with maximal $m = N - 1$ (implying $2N - 1 = 1/r$), i.e. with gaps of minimal length. For those fractals, relations (5) and (6) hold exactly for all R and the lacunarity is $F^{-1} = 1$ (minimal). The triadic Cantor set and the fractal of Fig. 1(a) are examples of this particular type. However, the triadic Cantor set is somewhat special: since $N = 2$, the only possible value of m is one and $F^{-1} = 1$ is its largest possible value. This explains why the gap is relatively large, despite the lacunarity being minimal. Nevertheless, we can construct fractals with its same dimension and larger lacunarity, for example, by taking $r = 1/9$ and $N = 4$; namely, the cases $m = 1, 2$. The case $m = 3$ gives rise to fractals with $F^{-1} = 1$, one of which is actually the triadic Cantor set.

3 SCALING OF VOIDS IN DIMENSIONS 2 AND 3

The definition of a void is simple and clear cut in dimension $d = 1$, because a point divides a segment into two disconnected parts. When dealing with point sets in higher dimensions, voids are usually ill defined since empty areas or volumes are (usually) connected. Indeed, the factor $c(r, N, m)$ in equation (3) was taking care of the connection between adjacent voids, and for $d > 1$ more than one definition is possible. A possible generalization of the definition in the previous section for $d > 1$ would be that only voids of equivalent size (that is to say, in a given iteration) are allowed to coalesce into a larger void. In this case, our results can be straightforwardly generalized and equation (5) reads

$$\Lambda(R) \approx \frac{1 - r^d N}{m} \left(\frac{N-1}{m} \right)^{-d/D} R^{-d/D}, \tag{9}$$

where Λ now stands for the area or the volume (in units of $1/r^d$) in dimension $d = 2, 3$, respectively. A particular generator is the one that starts with a square and removes a number of the $1/r^2$ square parts in a manner symmetrical with respect to a diagonal. The result is just a cartesian product of one-dimensional Cantor sets. The simplest example is obtained by taking $r = 1/3$ and removing five patches, forming a cross, out of the nine initial ones. The fractal so generated is the cartesian product of triadic Cantor sets. Clearly, the complementary set of these fractals is connected, but one can *assume* that the voids produced at every iteration are independent; then, one gets equation (9).

Other definitions of what constitutes a void are also possible. If we change our definition, the prefactor of $\Lambda(R)$ will change its precise form, and the estimated lacunarity of the fractal will also change accordingly. None the less, the scaling form of the Zipf law with void size remains unchanged, since it is independent of m and c. In the following, we show through numerical examples that any reasonable definition of void that is coherently applied at all scales returns the correct scaling for the Zipf law, and thus allows one to obtain quantitative information on the geometry of the distribution of points in the fractal set.

3.1 Voids in random fractals

In order to move towards the description of fractals arising in natural processes, we should first relax the deterministic character of their construction. Consider now a two-dimensional set for which $r = 1/3$ and $N = 4$, but where the five areas to be removed at each step in the construction are chosen at random. However, to mimic the observed morphology of the galaxy distribution, we must impose some constraints: the galaxy distribution has been characterized as a sponge-like network, with filaments and walls where galaxies accumulate (Gott, Melott & Dickinson 1986). In two dimensions, we should generate a fractal with some trace of filamentary structure. A rough way to achieve this is by constraining the five patches removed at every step to form a particular 'convex-like' shape – namely, a four-piece square with an extra piece adjacent to one side.

Randomization worsens the scaling range, so one should iterate the random generator many more times than the deterministic one to keep the scaling range similar. Alternatively, we will choose to average over independent realizations of the fractal constructed with the same number of iterations to improve the measures.

Since the problem of defining voids in two-dimensional sets is similar to that arising in three dimensions, we will work with $d = 2$ and extrapolate our results to higher dimensionality. Fig. 2 represents a fractal constructed by removing five randomly chosen (connected) patches of area r^{2i} at iteration i. Now, the possibility arises that voids produced in subsequent iterations are adjacent to previously existing ones and result in a more or less apparent increase in the size of a large void. As we pointed out, we wish to design an algorithm to find the voids in our set and apply it at all scales. Our working hypothesis is that the precise shape of the void is not relevant to recover the scaling behaviour of Zipf's law, as long as it is kept constant through iterations, and thus it is not relevant either to recover the fractal dimension of the associated distribution of matter.

Simple recursion relations of the previous type (Section 2) cannot be inferred when studying fractal sets arising from physical processes. Instead, one faces a set of points irregularly distributed in space and has to resort to other methods to estimate the distribution of voids. One of the simplest ways of defining an area devoid of points in the structure is the following.

(i) Coarse-grain your system by defining elementary cells such that there is at the most one point per cell (and defining so a lower cut-off to scaling, see Section 4).

(ii) Decide for regular elements to cover empty areas (say a square or a circle in $d = 2$).

(iii) Locate the largest square/circle centred at each empty cell (its boundary is limited by filled cells or sample boundaries).

(iv) Select the largest one, which is by definition a void of size Λ_1, and assign it rank $R = 1$.

(v) Fill the cells in the selected void (equivalent to those in the fractal set).

(vi) Repeat the procedure with the remaining empty cells until all of them are covered (this is somehow reminiscent of the box-counting method to estimate the fractal dimension Falconer 1990).

When the algorithm finishes, an ordered list of voids of decreasing size is produced. In Fig. 2(a) we represent an example of the

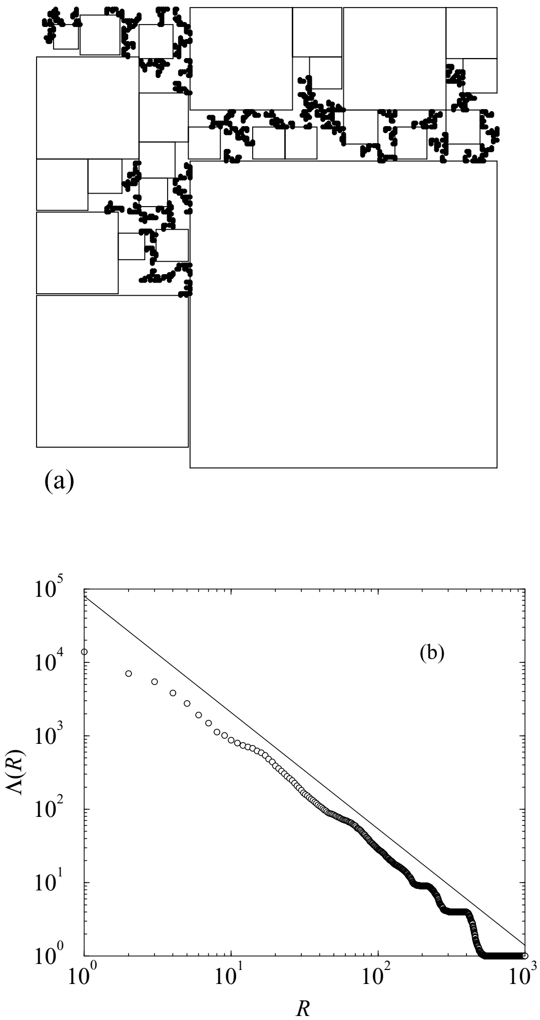

Figure 2. (a) Fractal point set generated by randomly removing five connected intervals of linear size $r = (1/3)^i$ at each iteration i. The squares correspond to the largest 25 voids found in the system through the algorithm described in the text. (b) Zipf's law for void sizes calculated according to the same algorithm. The solid line has slope $-2/D$, where $D = \log 5/\log 3$. The normalization of this line (its height at the origin) is arbitrary. The numerical data has been averaged over 500 independently-generated fractals.

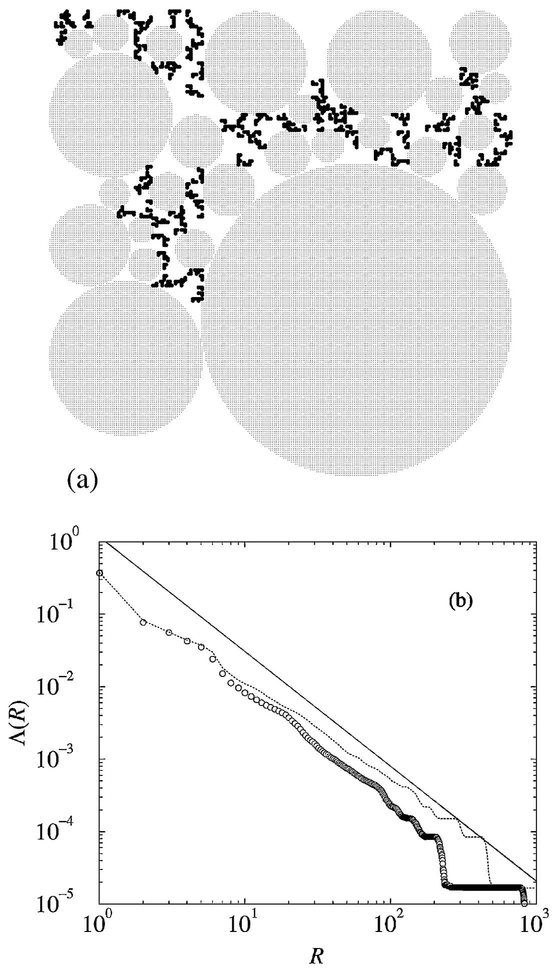

Figure 3. (a) The 25 largest circular voids found in the same set of Fig. 2 taking into account finite-size corrections. (b) Zipf's law. The dashed line shows $\Lambda(R)$ for circular voids calculated according to the algorithm described in Section 3. The circles show $\Lambda(R)$ for circular voids when only those with boundaries limited by the fractal set are counted. The scaling improves in the second case. Again, the solid line only indicates the expected slope, $-2/D$. Numerical results have been averaged over 100 independently-generated fractals.

first stages of the void-finding algorithm applied to a random fractal generated with the 'filamentary' generator, while Fig. 2(b) represents the obtained function $\Lambda(R)$ for square voids. As can be seen, the voids measured with the previous algorithm follow indeed the scaling expected according to the analytic prediction for deterministic fractals. We have applied our method to fractals constructed with several different generators in $d = 2$ and, in all cases, have obtained a good quantitative agreement between the predicted slope $-d/D$ and the numerically obtained one.

Another regular shape for voids that we have investigated is the circular one. Although largest voids clearly become even larger when a circular coverage is used, boundaries among voids define smaller voids which, however, are not limited by points in the fractal set (see Fig. 3). A first attempt to correct for these boundary effects would be to reject, in the final count, the voids that do not touch at least one point of the fractal set (that is, that only touch the external boundaries and/or other voids). This modification of our previously defined algorithm returns a better scaling for average sizes in the case of circular voids (both far from the boundary and from the elementary cell, see Fig. 3). These size classes were overloaded with 'artificial' voids placed among previously defined voids and/or external boundaries.

4 MEAN SIZE OF VOIDS IN A FRACTAL

As shown before, the distribution of voids in a fractal is a power law and, therefore, the mean size of voids is not a characteristic value, being dependent on the upper and lower cut-offs to the fractal scaling. Actually, the mean size of voids has been employed as a test for fractality, under the assumption that in a fractal the mean size of voids is proportional to the size of the sample, that is, the upper cut-off (Einasto et al. 1989). We shall show that this is not always the case: in fact, the mean size of voids is usually dependent on both cut-offs.

To calculate the mean size of voids we must use the probability distribution of sizes $p(\Lambda) \propto \Lambda^{-\alpha}$, with $\alpha = 1 + D/d$. Then,

$$\bar{\Lambda} = \frac{\int_{\Lambda_l}^{\Lambda_u} \Lambda p(\Lambda)\, \mathrm{d}\Lambda}{\int_{\Lambda_l}^{\Lambda_u} p(\Lambda)\, \mathrm{d}\Lambda}, \tag{10}$$

where Λ_l and Λ_u are the lower and upper cut-offs, respectively. The computation of the integrals is straightforward and

$$\bar{\Lambda} = \frac{-\alpha + 1}{-\alpha + 2} \frac{\Lambda_u^{-\alpha+2} - \Lambda_l^{-\alpha+2}}{\Lambda_u^{-\alpha+1} - \Lambda_l^{-\alpha+1}}. \tag{11}$$

For large Λ_u/Λ_l ratio (as required to have a reasonable scaling range) and taking into account that $1 < \alpha < 2$,

$$\bar{\Lambda} \approx \frac{\alpha - 1}{2 - \alpha} \Lambda_u^{2-\alpha} \Lambda_l^{\alpha-1}. \tag{12}$$

This expression cannot be reduced further and depends on both cut-offs. Note that if $D = d/2$ then $\alpha - 1 = 2 - \alpha = 1/2$ so that the mean size is just the geometric mean of both cut-offs (so to speak, both contribute equally); if $D > d/2$ then $\alpha - 1 > 2 - \alpha$ and the lower cut-off contributes more to $\bar{\Lambda}$, and vice versa if $D < d/2$.

We now discuss what values we should take for Λ_u and Λ_l in point distributions: whereas Λ_u is the well-determined size of the sample, the lower cut-off is trickier. Since a random fractal has a random component superposed on the deterministic algorithm that generates it, on the lowest scales the random component dominates and the distribution is approximately of Poisson type but with a very low mean number of points in the associated volume (*shot noise*). The crossover scale from the Poisson to a correlated (fractal) regime such that the number function[1] $\mathcal{N}(\mathcal{R}) = B\mathcal{R}^D$ is $B^{-1/D}$ (this scale has been defined by Balian & Schaeffer 1989 in a more general context). It is not totally clear what value is adequate for Λ_l in a galaxy catalogue. It must be larger than $(0.1\ \mathrm{Mpc})^3$ but it can be considerably larger, according to the way the catalogue has been compiled and the algorithm selected to find voids. At any rate, keeping Λ_l fixed one obtains that $\bar{\Lambda}$ is not proportional to Λ_u but rather to a power of it with exponent $2 - \alpha = 1 - D/d$ such that $0 < 1 - D/d < 1$ $(d = 3)$. A value of this exponent close to 1 (as reported by Einasto et al. 1989, see also Section 5) would imply a fractal dimension $D \ll 1$. A set with $D = 0$ is not really a fractal, but a collection of isolated points.

We have carried out numerical measurements on 2-dimensional fractals of known dimension D to test the accuracy to which a real fractal sample follows the scaling (12). Fig. 4 depicts some of our results. There, we have generated a fractal with $N = 3$ and similarity ratio $r = 1/2$, hence $D = 1.58$. The single patch to be removed at each iteration was chosen at random (see insert in Fig. 4). In order to see how $\bar{\Lambda}$ depends on Λ_u we have kept the lower cut-off fixed and equal to the size of the individual cell, $\Lambda_l = 1$ and varied Λ_u. As long as $\Lambda_l \ll \Lambda_u$ we observe a neat scaling with the predicted exponent $2 - \alpha$. Next, the upper cut-off was kept fixed to its maximum value for this fractal set, $\Lambda_u = 2^{14}$. The variation of Λ_l was carried out in practice by averaging only over voids of area equal to or larger than Λ_l. In this second case, the asymptotic scaling exponent is $\alpha - 1$. Numerical results are compared with the two curves obtained from equation (11) with $\Lambda_l = 1$ and $\Lambda_u = 2^{14}$, respectively. Only when both cut-offs become of comparable order are deviations from scaling observed.

[1] The number function is the mean number of particles inside a ball of radius $\mathcal{R}$ centred on a particle and equals the mass function $\mathcal{M}(\mathcal{R})$ divided by the mass of a particle.

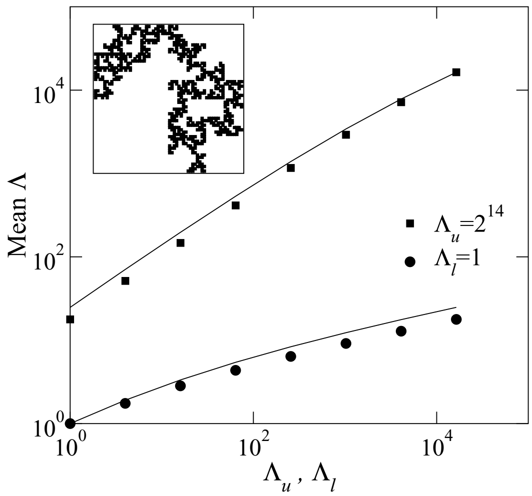

Figure 4. Scaling of the mean size of voids when one cut-off is varied and the other kept fixed. The insert depicts a random fractal with $N = 3$ and $r = 1/2$ embedded in a space of dimension $d = 2$. Solid lines correspond to a numerical solution of the expression for $\bar{\Lambda}(11)$ corresponding to this fractal (with $D = 1.58$). The slight difference between numerical and analytical results is due to the substitution of a sum (over discrete void areas) by an integral in the expression for $\bar{\Lambda}(10)$.

5 GALACTIC VOIDS

Galactic voids are vast regions of space apparently devoid of luminous matter (galaxies). Several authors have developed algorithms in order to detect the extent and frequency of such regions in current galaxy catalogues (Einasto et al. 1989; Kauffmann & Fairall 1991; Hoyle & Vogeley 2002). The aim of these studies is to gain a better understanding of the morphology of the distribution of galaxies in order to, eventually, correlate it with the physical mechanisms responsible for the observed structure. A first step in this direction has been the comparison of measures of voids made on the current galaxy catalogues with measures on *N*-body simulations of cold dark matter (Müller et al. 2000; Arbabi-Bidgoli & Müller 2002). The problem of defining what constitutes a void in $d = 3$ has been ubiquitous, and all these studies have solved this indeterminacy in different ways. Interestingly, all of them have looked for maximal volumes of approximately convex shape (but differently shaped depending on the area studied) inscribed among matter points, while none has considered regular volumes. Our main point here is that, when trying to recover quantitative information, shape matters, implying that voids have to maintain their shape at all scales. On the one hand, we have shown that this criterium permits us to obtain information on a fractal distribution of matter. On the other hand, our definition eliminates certain arbitrariness in void finding algorithms, for instance the amount of overlap between voids to be merged into a single larger void.

With these caveats in mind, we have examined two studies reporting large voids and examined the function $\Lambda(R)$ that they produce. Recently, Hoyle & Vogeley (2002) have examined the Point Source Catalogue Survey (PSCz) and the Updated Zwicky Catalog (UZC) for the presence of voids. The volume of the 35 and 19 largest voids (respectively) is plotted in Fig. 5 attending to their rank. Although, in these cases, $\Lambda(R)$ is relatively well fitted by a straight line for the largest voids, the slope is too low to represent the complementary set (that is, the set of voids) of a fractal distribution of matter. There is a physical restriction to the exponent of the scaling law, since the dimension of the fractal cannot be

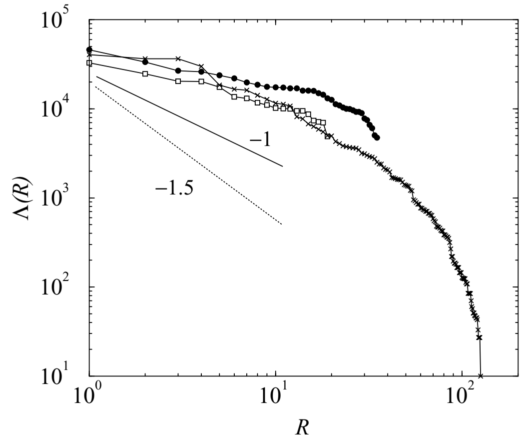

Figure 5. Zipf's plot for three available void catalogues. Data represented as circles (PSCz) and squares (UZC) are from Hoyle & Vogeley (2002), units are h^{-3} Mpc^3; data in crosses from Kauffmann & Fairall (1991), units are km^3 s^{-3}. Under the hypothesis that matter is self-similarly distributed in the universe, the scaling expected for void volumes is $\Lambda(R) \sim R^{-3/D}$, where D stands for the fractal dimension of the galaxy distribution. The solid line has slope -1 (indeed, any function $\Lambda(R)$ has to have slope larger than unity in absolute value). The dotted line signals the expected scaling if, according to recent measures, $D \simeq 2$, leading to the -1.5 scaling relation shown in the figure. The dotted and the solid lines just indicate the (expected) slope of $\Lambda(R)$ in each case.

larger than that of the embedding space, i.e. $d > D$. This implies that the exponent $-d/D$ has to be larger than unity in absolute value. We represent this restriction in Fig. 5 and observe that the putative slopes are much lower than this value. The classical value of the fractal dimension, deduced from the two-point correlation function $\xi(r) \propto r^{-\gamma}$, is $D = 3 - \gamma \simeq 3 - 1.8 = 1.2$. However, recent reanalyses of the galaxy catalogues (Sylos Labini et al. 1998) yield a larger value – namely, $D \simeq 3 - 1.1 = 1.9$. In this case, the expected exponent for $\Lambda(R)$ versus R would be about -1.5, which we represent as a dotted line in Fig. 5.

Ten years ago, Kauffmann & Fairall (1991) developed an algorithm to search for voids. They reported a list of 129 'significant' voids obtained from the merged Southern Redshifts Catalogue and the Catalogue of Radial Velocities of Galaxies. Their list is represented in Fig. 5 together with the previous data. Apart from an initial almost flat stage, we have found that this function is reasonably well fitted by an exponential law (not shown in the figure). This does not correspond to a fractal geometry and would rather correspond to a Poisson distribution of points, where voids much larger than the volume per point should be exponentially suppressed. However, they define a *significant void* as one that '. . . occurs in the random catalogue simulations with probability less than 1 per cent' (see Kauffmann & Fairall 1991 for more details). Moreover, they take as the reference Poisson distribution for a given catalogue the one with the same number of points but randomly distributed. While one certainly cannot consider voids as significant below the scale of the lower cut-off Λ_1, where the distribution can be effectively considered Poissonian (as remarked in Section 4), their procedure produces a Poissonian distribution with a much larger mean interparticle distance. Hence, even though very large voids are almost always significant in their sense, this is not the case for average- and small-size voids, which occur frequently in a random catalogue with the same number of points. Those voids are not included in the list they provide, and hence the distribution is strongly depleted in the mean and small-void domains.

Einasto and co-workers (1989) have shown that mean void diameters increase with the sample size in a power-law manner. They have identified this fact as indicative of self-similarity in the matter distribution. Their qualitative results agree with the calculations reported here (Section 4). However, the quantitative result does not agree with our prediction. In the work by Einasto et al. (1989), we understand that measurements were carried out in such a way that the lower cut-off was kept fixed, while the upper cut-off varied. One expects then a scaling of the form $\bar{\Lambda} \propto \Lambda_u^{\beta_u}$, with $\beta_u = 1 - D/d$. Since $0 \leqslant D \leqslant d$, the exponent is bounded, $0 \leqslant \beta_u \leqslant 1$, with $\beta_u = 0$ corresponding to a homogeneous distribution of points and $\beta_u = 1$ to an (almost) empty set of points. This second value is the one reported by Einasto et al. (1989).

All our considerations apply to galaxies as particles with no features – in other words, ignoring type and luminosity. It has been argued that voids could be populated by faint galaxies, called *field galaxies* to distinguish them from 'normal' cluster galaxies (El-Ad et al. 1997; Hoyle & Vogeley 2002). Luminosity can be taken into account by generalizing the concept of a fractal distribution to a multifractal distribution (Sylos Labini et al. 1998), which we do not consider here. If the neglected field galaxies have a distribution with geometrical properties different from the more luminous ones, the multifractal model would not apply. However, the standard biased galaxy formation picture attributes similar scaling properties to the distributions at various biasings (Gabrielli, Sylos Labini & Durrer 2000), in accordance with a multifractal distribution. A sort of biasing could be mimicked for a pure fractal by randomly removing a fraction of points, which would not alter its scaling properties. At any rate, a thorough analysis of this question falls beyond the scope of this work.

5.1 Sources of deviation from scaling

Apart from our previous discussion on the way in which voids are defined and counted, there are a number of mechanisms which, in our understanding, could produce a significant deviation from the scaling regime. We list them and briefly discuss their effects. Sometimes the source of deviations can be identified and even eliminated. But often, even in what sense they would affect Zipf's plot is unclear. The following list might not be exhaustive, but some or even all of the listed problems can affect the observations to date. However, note that scaling corresponding to $D > d$ should not occur, since it is physically forbidden.

(i) *Finite size effects.* Usually, it is unavoidable to use a 'boundary' to limit the fractal set that we are measuring. We have already seen that, in particular for certain shapes of voids, systematic deviations from scaling can be obtained.

(ii) *Scale-dependent dimension of galaxy distribution.* It has been recently reported (Bak & Chen 2001) that the dimension of the galaxy distribution varies with the observation scale. It grows smoothly from zero when approaching the size of single galaxies to three at the scale where the transition to homogeneity[2] takes place. Since voids involved in Zipf's plot would cover the whole range of sizes, there might be some systematic deviations in cases where the fractal dimension is scale dependent: D decreases with decreasing scale, hence the exponent of the rank-ordering plot increases in absolute value, and the function becomes concave from below.

[2] The concept of homogeneity involves some subtleties (Gaite, Domínguez & Pérez-Mercader 1999). Here and henceforth, we mean by homogeneity that the relative density fluctuations are small.

(iii) *Transition to homogeneity.* There cannot be voids larger than the characteristic length at which the universe becomes homogeneous; but the characteristic size of the largest voids is an independent scale (Balian & Schaeffer 1989) that could be significantly smaller than the homogeneity scale. Furthermore, the breakdown of scaling in the void distribution at the characteristic size of the largest voids might suggest that the recent observations reported by Hoyle & Vogeley (2002) returning a flat $\Lambda(R)$ are related to it.

6 CONCLUSIONS

There is a quantitative and well-defined relationship between the dimension D of a fractal set and the exponent of Zipf's plot $\Lambda(R)$ for the corresponding void sizes Λ. We have illustrated this dependency with regular fractals in $d = 1$, for which exact relations have been derived. Next, the introduction of a simple algorithm to identify voids in any dimension has allowed us to show that the relation $\Lambda(R) \simeq R^{-d/D}$ also holds in stochastic fractals defined in dimension $d = 2$. We have shown that the mean size of voids in a sample defined between a lower and a higher cut-off scales with these quantities, $\bar{\Lambda} \propto \Lambda_u^{\beta_u} \Lambda_l^{\beta_l}$. The exponents β_u and β_l depend on the fractal dimension D; hence, the relation between $\bar{\Lambda}$ and the two cut-offs depends on the fractal. Our results can be straightforwardly extrapolated to $d = 3$.

This study has been performed with the aim of applying it to current measures of the distribution of galaxies. On the one hand, current measures of the two-point correlation function seem to be consistent with a fractal distribution, with a yet uncertain dimension $1 < D < 2$, and in a still controversial range from 1 to, perhaps, $\sim 100\ h^{-1}$ Mpc (or even more) (Guzzo 1997; Sylos Labini et al. 1998; Martínez 1999). On the other hand, current void catalogues (Kauffmann & Fairall 1991; Hoyle & Vogeley 2002) do not seem to support this result. None the less, attending the discussion which conforms the body of our work, they are insufficient to discard the hypothesis of a fractal distribution of galaxies. To assess the capability of void finding algorithms to detect fractal structure and then the fractal dimension D, we would suggest that they be tested with simple examples, for which exact results can be easily obtained, as we have done here.

It would be interesting to extend the measures of void sizes to scales smaller than the ones usually probed in a systematic way and, specifically, compare them with a scaling distribution. The careful construction of the Zipf plot of void sizes and the analysis of its scaling (or not), as well as the scaling of the average size of voids with the sample size are complementary measures to n-point correlation functions and additional support for the fractality of the distribution of galaxies. In any case, it is clear that the two methods must provide equivalent information, meaning that the study of the convergence of both approaches could help distinguish different sources of deviation from scaling and moreover better characterize the morphology of the galaxy distribution.

ACKNOWLEDGMENTS

We thank Ugo Bastolla, Alvaro Domínguez, Antonino Giaquinta and Juan Pérez-Mercader for interesting discussions and support. The work of both authors is currently supported by respective Ramón y Cajal contracts of the Ministerio de Ciencia y Tecnología.

REFERENCES

Arbabi-Bidgoli S., Müller V., 2002, MNRAS, 332, 205
Bak P., Chen K., 2001, Phys. Rev. Lett., 86, 4215
Balian R., Schaeffer R., 1989, A&A, 220, 1
Blumenfeld R., Ball R. C., 1993, Phys. Rev. E, 47, 2298
Chown M., 1999, New Scientist, 2200, August 21
Coleman P., Pietronero L., 1992, Phys. Rep., 213, 311
Einasto J., Einasto M., Gramann M., 1989, MNRAS, 238, 155
Einasto J., Jõeveer M., Saar E., 1980, MNRAS, 193, 353
El-Ad H., Piran T., da Costa L. N., 1997, MNRAS, 287, 790
Falconer K., 1990, Fractal Geometry. Wiley, Chichester, UK
Gabrielli A., Sylos Labini F., Durrer R., 2000, ApJ, 531, L1
Gaite J., Domínguez A., Pérez-Mercader J., 1999, ApJ, 522, L5
Gott J. R., III, Melott A. L., Dickinson M., 1986, ApJ, 306, 341
Guzzo L., 1997, New Astronomy, 2, 517
Hoyle F., Vogeley M. S., 2002, ApJ, 566, 641
Kauffmann G., Fairall A. P., 1991, MNRAS, 248, 313
Mandelbrot B. B., 1977, The Fractal Geometry of Nature. W. H. Freeman, New York
Martínez V., 1999, Sci., 284, 445
Müller V., Arbabi-Bidgoli S., Einasto J., Tucker D., 2000, MNRAS, 318, 280
Peebles P. E. J., 1980, The Large-Scale Structure of the Universe. Princeton Univ., Princeton, NJ
Sornette D., 2000, Critical Phenomena in Natural Sciences. Springer-Verlag
Sylos Labini F., Montuori M., Pietronero L., 1998, Phys. Rep., 293, 61
Vogeley M., Geller M. J., Huchra J. P., 1991, ApJ, 382, 44
Wu K. K. S., Lahav O., Rees M. J., 1999, Nat, 397, 225
Zipf G. K., 1949, Human Behavior and the Principle of Least Effort. Addison-Wesley, Cambridge, MA

APPENDIX A: ZIPF'S LAW FOR A DETERMINISTIC FRACTAL

We show here that the function $\Lambda(R)$ has (approximately) equal steps in logarithmic scale and how to eliminate the discrete variable i from equations (3) and (4).

We can asymptotically approximate equation (4) by

$$R_i = 1 - \frac{m}{N-1} + m\frac{N^{i-1}}{N-1} \approx m\frac{N^{i-1}}{N-1} \tag{A1}$$

for large R_i (which implies that i is large, assuming that N and m are not). The step length in logarithmic scale is

$$\ln\left(R_i + mN^{i-1}\right) - \ln R_i = \ln\left(1 + \frac{mN^{i-1}}{R_i}\right) \approx \ln N. \tag{A2}$$

On the other hand, within the same approximation, after taking logarithms of equations (3) and (4),

$$\ln \Lambda_i = \ln c + i \ln r, \tag{A3}$$

$$\ln R_i = \ln \frac{m}{N-1} + (i-1)\ln N. \tag{A4}$$

Now, it is easy to solve for i in the second equation and substitute it in the first one, obtaining

$$\begin{aligned}\ln \Lambda_i &= \ln r\left(\frac{\ln R_i - \ln[m/(N-1)]}{\ln N} + 1\right) + \ln c \\ &= \frac{-1}{D}\left(\ln R_i + \ln\frac{N-1}{m}\right) + \ln r + \ln c.\end{aligned} \tag{A5}$$

From $cm + N = 1/r$, $rc = (1 - rN)/m$. Then, after removing the logarithms,

$$\Lambda_i = \left(\frac{N-1}{m}\right)^{-1/D} \frac{1-rN}{m} R_i^{-1/D}. \tag{A6}$$

Let us briefly analyse the accuracy of the approximation (A1). For the example in Fig. 1, with $N = 5$ and $m = 1, 2, 4$, $R_3 = 7, 13, 25$, whereas the approximation yields 6.25, 12.5, 25. Of course, the accuracy is higher for $i > 3$. In general, the relative error is $O(N^{-i})$.

This paper has been typeset from a TEX/LATEX file prepared by the author.