Mon. Not. R. astr. Soc. (1985) **216,** 17–35

Minimal spanning trees, filaments and galaxy clustering

John D. Barrow[1], Suketu P. Bhavsar[1,2] and D. H. Sonoda[1]

[1] *Astronomy Centre, University of Sussex, Brighton BN1 9QH*

[2] *Raman Research Institute, Bangalore 560080, India*

Accepted 1985 March 11. Received 1985 February 28; in original form 1985 February 14

Summary. We describe a graph theoretical technique for assessing intrinsic patterns in point data sets. A unique construction, the minimal spanning tree, can be associated with any point data set given all the inter-point separations. This construction enables the skeletal pattern of galaxy clustering to be singled out in quantitative fashion and differs from other statistics applied to these data sets. We describe and apply this technique to two- and three-dimensional distributions of galaxies and also to comparable random samples and numerical simulations. The observed CfA and Zwicky data exhibit characteristic distributions of edge-lengths in their minimal spanning trees which are distinct from those found in random samples. These statistics are also re-evaluated after normalizing to account for the level of clustering in the samples.

1 Introduction

The availability of the CfA redshift survey (Davis *et al.* 1983) has provoked theorists to evaluate the statistical significance of apparent filamentary features in the observed two- and three-dimensional distribution of galaxies (Kuhn & Uson 1982; Einasto *et al.* 1982; Bhavsar & Barrow 1983, 1984; Dekel & West 1985). Traditional statistical measures, like the two-point correlation function introduced by Totsuji & Kihara (1969) and used extensively by Peebles (1980), are not sensitive to specific patterns or linear features in the luminous matter distribution and must be supplemented by additional statistics that more effectively measure the extent of intrinsic one- and two-dimensional features. This question is of some importance because it has been suggested, but not rigorously demonstrated, that the 'pancake' and 'hierarchical' theories of galaxy formation might produce very different degrees of filamentarity in the distribution of visible galaxies (Zeldovich, Einasto & Shandarin 1982). A number of simple geometric ideas have been tried in order to quantify linear patterns in two-dimensional data sets (Kuhn & Uson 1982; Moody, Turner & Gott 1983; Dekel & Aarseth 1984) and percolation analyses have been performed on various observational samples and numerical simulations of clustering (Bhavsar & Barrow 1983, 1984; Einasto *et al.* 1982; Dekel & West 1985). So far, none of these measures

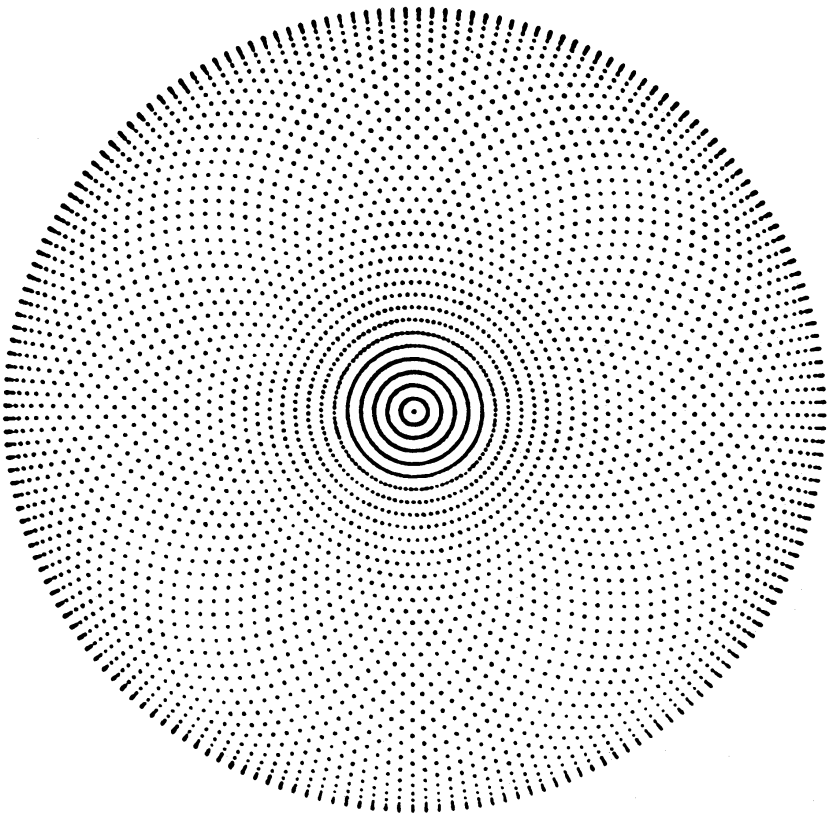

Figure 1. The recognition of pattern by the human eye. The eye perceives the intrinsic point pattern as a set of concentric circles near the centre of the array but towards the outside the predominant effect is that of petal-like curves. This arises because the eye reinforces the nearest-neighbour relationships. Near the centre the nearest neighbours to any point lie on the same circle but towards the outside they lie on adjacent circles of different radius. If one looks now at this same picture along the plane of the page, then the perceived patterns alter dramatically because the eye now perceives the nearest neighbour distances *in projection*. The 'real' patterns intrinsic to this array are concentric circles and radial straight lines through the centre but both are disguised by the pattern recognition technique of the human eye. The overlaying of a transparency of the same array produces further unusual patterns.

appears to be a compelling discriminator of filamentary structure, able to distinguish between alternate theoretical models. Here, we introduce and illustrate the use of a technique which employs concepts borrowed from graph theory (Ore 1962). It provides a quantitative measure of intrinsic pattern within a point data set. It can also be employed as a filament-finding algorithm.

It is illuminating to carry out an experiment which illustrates the way in which the human eye's perception of pattern is dominated by nearest neighbour effects. The pattern shown in Fig. 1 is simply an array of concentric rings of dots. Each ring contains the same number of dots and is displaced relative to adjacent rings so that dots on every other ring lie on radial straight lines through the centre of the ring system. Thus the built-in pattern is a mixture of circles and radial lines. However, if we look at the figure we notice that the eye traces circles near the centre but petal-like patterns towards the outside. These are the patterns which trace out the nearest neighbour relationships between points. If one looks at the figure along the plane of the page the perceived patterns are changed completely into a nested array of 'figure eights'. Any slight bending of the page during this exercise can create other, different apparent patterns. In this case, when looking along the edge of the page, the inter-point distances are viewed in projection and so are changed, hence the eye is led to pick out a new dominant pattern.

This experiment shows how subjective pattern recognition can be and how sensitive the

impression obtained by the eye actually is to minor perturbations which affect nearest neighbour distances. It is worth noting that so far, the presence of filamentary structure in the distribution of galaxies is a subjective visual impression only.

2 Minimal spanning trees

Any distribution of points in space has associated with it a unique network (Gower & Ross 1969; Abraham 1962; Zahn 1971), called its *minimal spanning tree* (MST), which picks out the dominant pattern of connectedness in a manner which emphasizes its intrinsic linear associations. This enables the skeletal pattern within a point data set to be singled out in a systematic and repeatable fashion. Quantitative properties of the MST can be calculated for comparative purposes and used to define randomness in a manner that differs from other criteria used in previous studies of galaxy clustering.

Our data set is termed a *graph* (Iyanaga & Kawada 1980) and will be composed of a collection of *nodes* (galaxies), *edges* (straight lines joining galaxies) and *edge-lengths* (distances between galaxies). A sequence of edges joining nodes is a *path*; a closed path is called a *circuit* and a graph will be called *connected* if there is a path between any pair of nodes. The number of edges emanating from a node is called its *degree*. A connected graph containing no circuits is a *tree*. If the tree of a connected graph, Γ, contains all the nodes of the data set then it is called a *spanning tree*. The length of a tree is defined to be the linear sum of the lengths of its component edges. The *minimal spanning tree* (MST) of Γ is the spanning tree of minimum length. If no two edge-lengths are equal the MST of Γ will be unique. A *k-branch* is a path of k edges connecting a node of degree 1 to a node of degree exceeding 2 with all intervening nodes of degree 2.

The simplest algorithm to construct explicitly the MST of a graph, Γ, first picks an arbitrary node of Γ and then adds the connected edge of smallest length. This edge and the two nodes at its ends form the *partial tree*, Π_1. The kth partial tree, Π_k, is formed by adding to Π_{k-1} the shortest edge connecting Π_{k-1} to any nodes of Γ not already in Π_{k-1}. If Γ contains n nodes then Π_{n-1} is the required MST. The construction of the MST is closely related to the formulation of the well-known 'travelling salesman' problem (Kruskal 1956), and our algorithm is based upon the fact that any sub-tree of the MST of Γ will itself be an MST for the graph formed by the nodes of the sub-tree.

The procedure described above picks out the underlying shape of the data set in a manner that is in accord with the eye's response (it is interesting to notice the correspondence between MST's and the lines that the ancients drew between the stars to create the graphs we call the constellations).

In order to distil off the dominant features of the MST above 'noise' it is sometimes advantageous to perform two further reducing operations (Clark & Miller 1966; Zahn 1971) on the MST:

(i) *Pruning:* An MST is pruned to level p when all k-branches with $k \leqslant p$ have been removed.

(ii) *Separating:* Remove from the MST any edges whose length exceeds some cut-off. For example, if $l > 3\langle l \rangle$ where $\langle l \rangle$ is the mean length of paths in the MST.

Operation (i) removes from the MST superficial 'foliage' which plays no role in delineating the principal pattern. Operation (ii) enables distinct subclusters to be isolated and 'accidental' long linkages arising from low point-densities to be eliminated. It also enables 'field' clumps to be identified. The skeletal MST that results after pruning and/or separating may consist of several disjoint pieces and can be subjected to various quantitative analyses. In Fig. 2 the process of MST construction followed by operations (i) and (ii) is illustrated by a simple two-dimensional example.

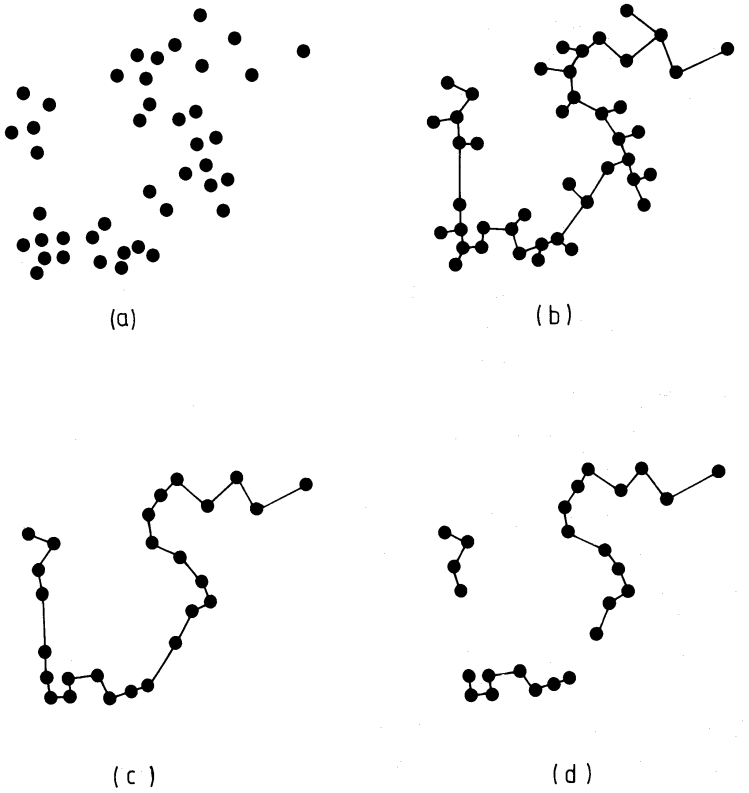

Figure 2. The construction and reduction of a minimal spanning tree (MST). (a) A point data set. (b) MST constructed according to the prescription given in Section 2 of the text. (c) Pruned MST; all nodes in (b) of degree one connected to nodes with degree exceeding two have been removed along with their connecting edges. (d) Separated and pruned MST; all edges of (c) exceeding a critical length have been removed.

One of the great advantages of the MST is that it selects a well-defined subset of all the possible connections between nodes. For a set of n modes, $n(n-1)/2$ different separations are possible. The MST will pick out a unique set of $(n-1)$ edge-lengths (although the actual sequence of connected nodes in an MST is not unique unless all the internode distances are different, the distribution of edge-lengths is). This, together with its ability to map out the main features of a graph, suggests that MST analysis may be well suited to the study of filamentary structure in the distribution of galaxies.

3 Applications in two dimensions

In order to illustrate the effectiveness of MST analysis we shall first apply the method to some random distributions and real galaxy catalogues in two dimensions where the results are easy to visualize. Our graphs consist of galaxies as nodes and angular separations as edge-lengths. As there appears to exist no general theory as to the expected structure of MST's (the work of Erdös & Renyi (1960) on random graphs does not investigate MST's), these experiments will allow us to build up a picture of the structure and robustness of MST's. The MST's we display have been generated by computer using the construction algorithm discussed above. Since the MST is unique in practice, the constructive process can begin at any point of the data set. In this paper we shall be interested principally in one quantitative property of the MST: *the fraction, F(l), of edges of length l in the MST.* The lengths, l, are scaled by the mean edge-length of all edges in the MST, $\langle l \rangle$. This gives a measure of pairwise correlations within a specially selected subset that uniquely traces the intrinsic pattern of the distribution.

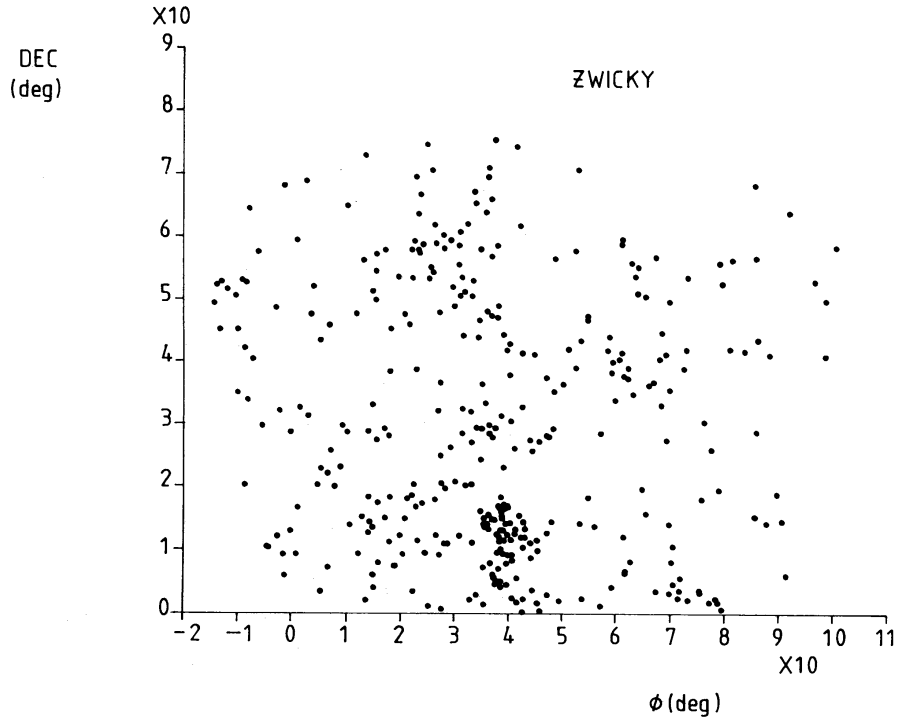

Fig. 3 (a)

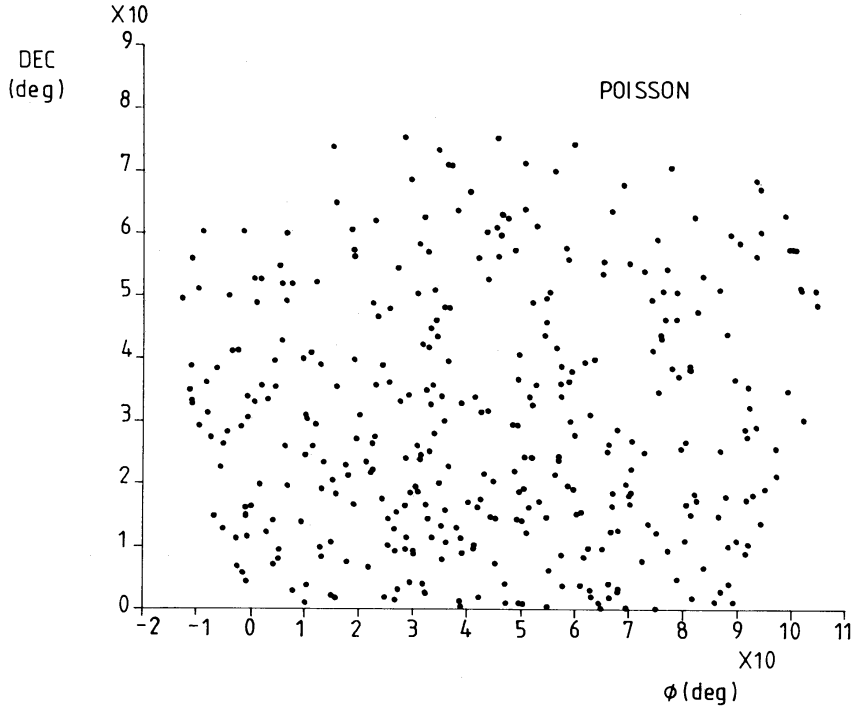

Fig. 3 (b)

Figure 3. Two-dimensional samples and minimal spanning trees. The angular positions of: (a) all 1091 Zwicky galaxies and (b) the equivalent random sample. 'Dec' is the delination in the equatorial coordinates system and ϕ is an azimuthal angle. The North Galactic Pole is at Dec=22.6°, ϕ=45°. The resulting MSTs, pruned to level 10 (i.e. all k-branches with $k \leqslant 10$ removed) are shown (c) Zwicky sample MST, and (d) the random sample MST. The mean edge-lengths used in each case were $\langle l_z \rangle$=0.0215 rad and $\langle l_r \rangle$=0.0267 rad; (e) the pruned and separated Zwicky MST, (f) the pruned and separated random MST. The cut-offs used in (e) and (f) are $2\langle l_z \rangle$ and $1.6\langle l_r \rangle$ respectively and correspond to the same absolute angular separation of 0.043 rad. It should be noted that Figs 3(a) and (b) are not conformal mappings of the celestial sphere and distortions occur at large declinations (for example at Dec=90° a point is mapped into a line).

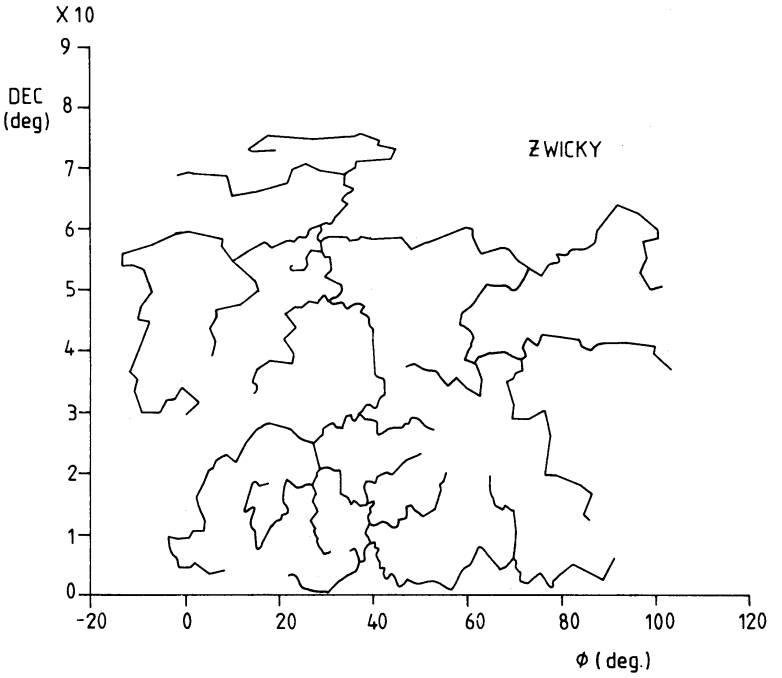

Fig. 3 (c)

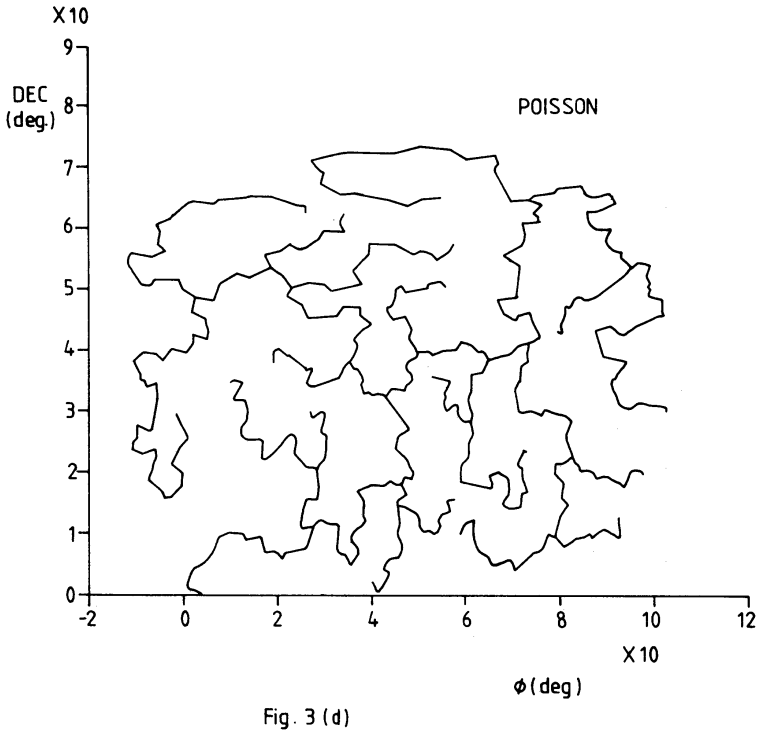

Fig. 3 (d)

Figure 3 – *continued*

In Fig. 3 we display 1091 galaxies in the North Galactic Cap defined by $\delta \geq 0°$ and $b^{II} \geq 40°$ from the Zwicky 14 mag catalogue (Zwicky *et al.* 1961–68) and a Poisson data set of 1091 galaxies occupying the same sky area. In each case we have also illustrated the full MST and the pruned and separated MST. Pruning was performed to level 10 (that is, all k-branches with $k \leqslant 10$ were removed) and the cut-offs used in separating were measured in units of the mean edge-length, $\langle l \rangle$. Since $\langle l \rangle$ differs for the Zwicky (z) and Poisson (r) samples, the cut-offs used, at $2\langle l \rangle_z$ and $1.6\langle l \rangle_r$

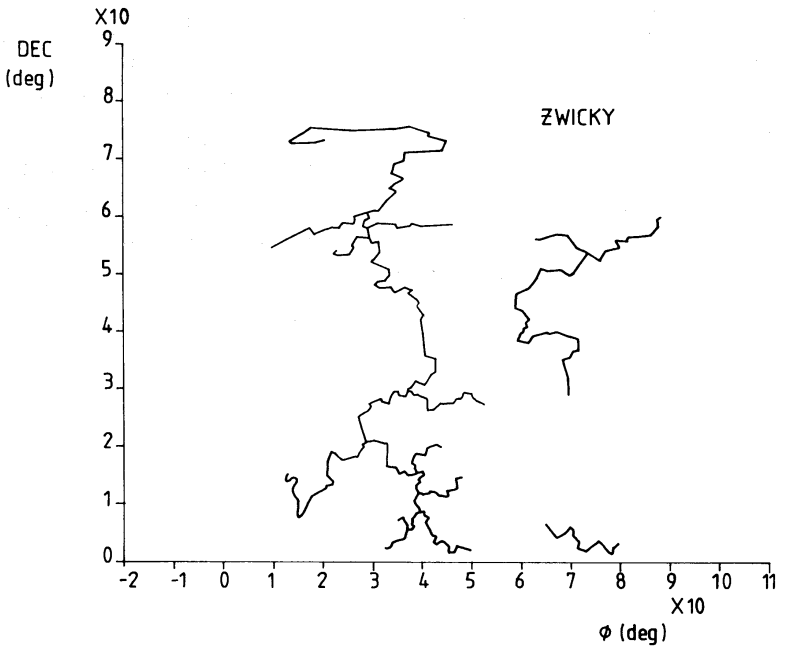

Fig. 3 (e)

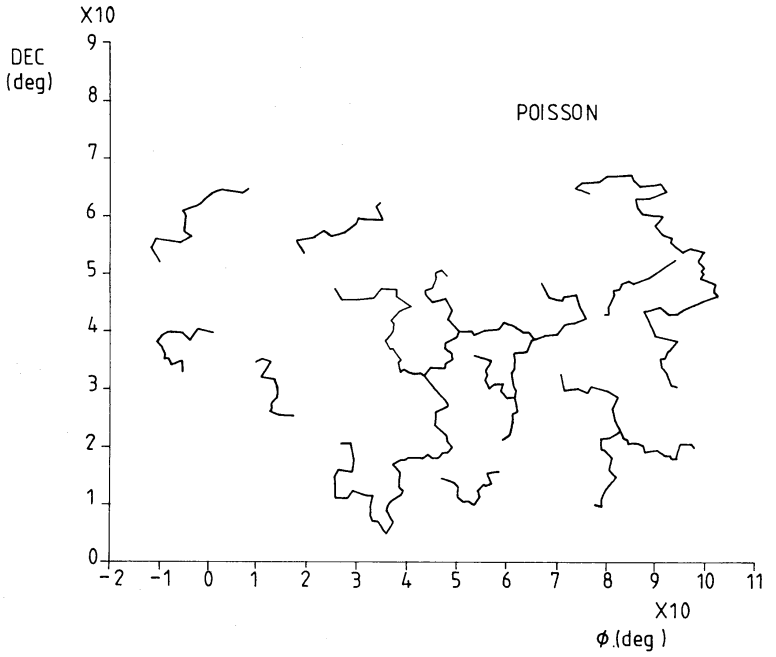

Fig. 3 (f)

Figure 3 – *continued*

for the Zwicky and Poisson samples respectively, were chosen so that they correspond to the same absolute angular separation of 0.043 radians in each case. Changing the value of the cut-off scale is similar to changing the radius of the spheres around each node in a percolation analysis (Bhavsar & Barrow 1983, 1984).

The Zwicky and random MST's have a quite different structure. The Zwicky MST has distinct directional bias and filamentary structure, contains few disjoint pieces and free ends, whereas the random MST exhibits no directional bias and contains a large number of small loops with many

free ends. By examining Fig. 3 one can see roughly how many branches would be removed from the MST when the cut-off length is varied. When separated, the MST of the Poisson distribution breaks up into many more disjoint pieces than the Zwicky MST. These differences reflect the purely visual differences between the two data sets. The Zwicky MST also picks out a dominant linear feature at $\phi \sim 40°$ whereas the Poisson MST exhibits no significant deviations from isotropy. Fig. 3(a) and (b) are not conformal mappings of the celestial sphere. This alters the appearance of trees near Dec=90°. In order to evaluate the ability of the MST operation to identify low-density filamentary features in a data set we have added a linear feature containing 20 points to the random data of Fig. 3. The resulting MST is shown in Fig. 4.

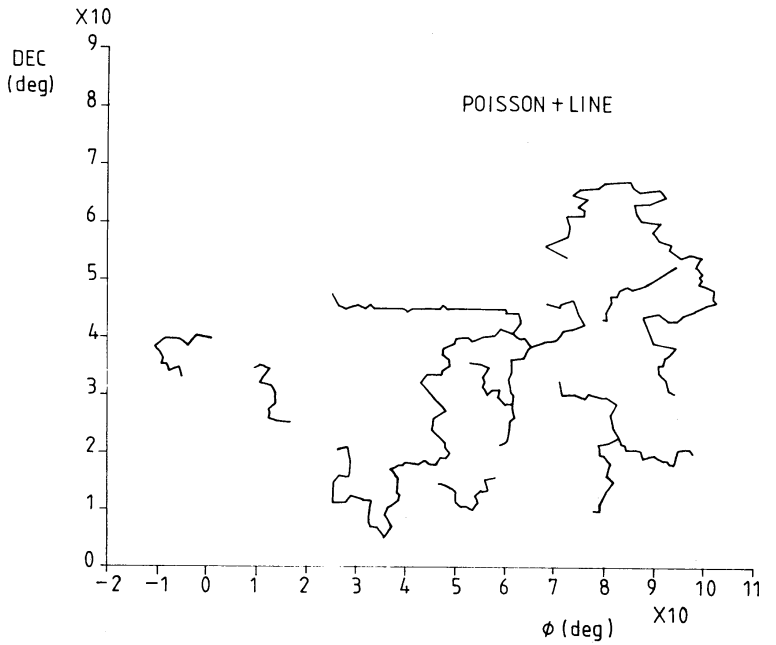

Figure 4. Detection of a linear filament. The effect on the MST shown in Fig. 3(d) of adding a linear feature of 20 points to the random point distribution shown in Fig. 3(a). Whereas such a small perturbation would have a negligible effect upon averaged statistics like the two-point correlation function of the sample, it is clearly highlighted by the MST construction.

The MST clearly picks out the linear feature we have superimposed and remains virtually unperturbed elsewhere. In general, the MST of any data set will be unchanged by any perturbation that conserves the ordering of the edge-lengths; in particular, this includes translations, rotations and dilations.

We have examined the number of nodes of a given degree to see if there were significant differences in the number of free ends (nodes of degree 1) or in the number of 'linear branches' (if there are long linear sections with few side-branches then we expect more nodes of degree 2). However, there were no striking differences between Zwicky and Poisson cases. The effect of pruning on the distribution of node degree was also the same for both types of sample.

Because of the convoluted appearance of the branches in the Poisson MST's, we also looked at the distribution of branch lengths and end-to-end separations. We define the *branch length* to be the piecewise sum of the edge-lengths along any path, whereas the *end-to-end* separation is the shortest distance between the first and last nodes of a branch. By comparing these two quantities for a given branch of the MST it was hoped that we could identify long linear branches (branch

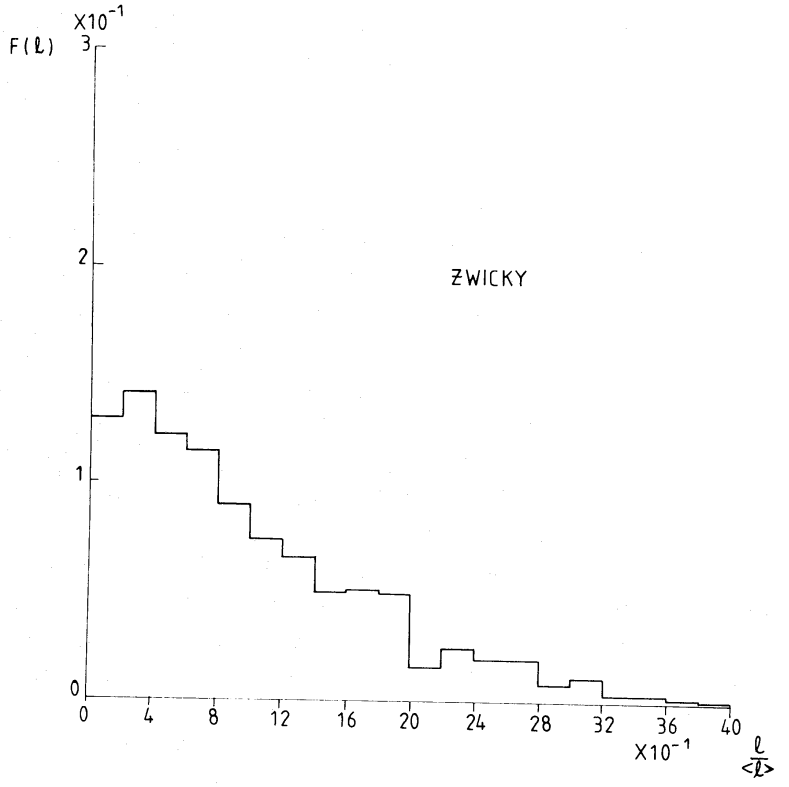

Fig. 5 (a)

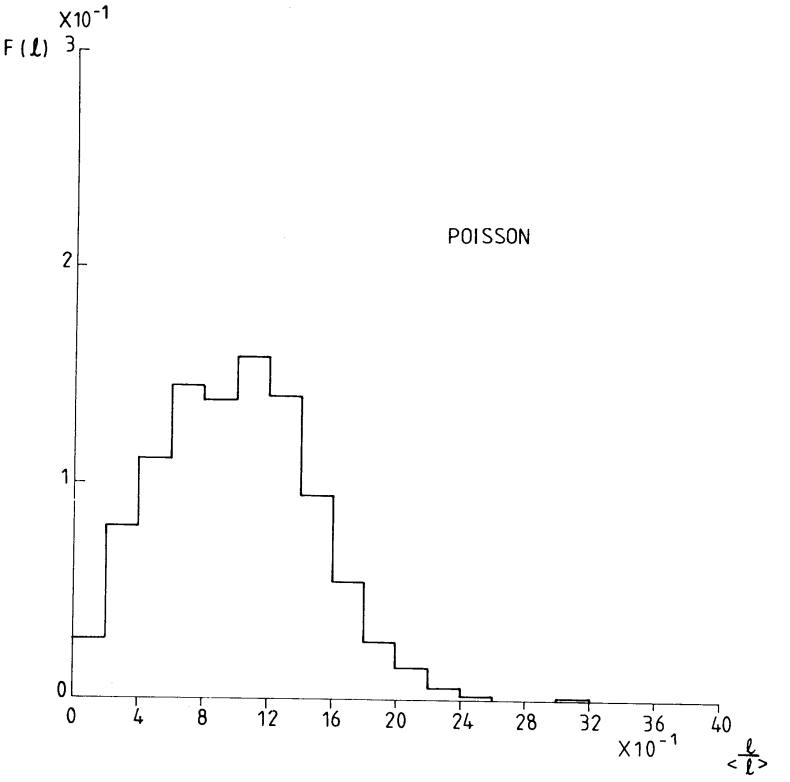

Fig. 5 (b)

Figure 5. Frequency distribution of two-dimensional MST edge-lengths. The $F(l)$ versus $l/\langle l\rangle$ distributions for the full MSTs of the (a) Zwicky and (b) random data sets displayed in Fig. 3. The $F(l)$ distribution for the random sample is well described by a Gaussian centred on $\langle l\rangle$. The Zwicky distribution exhibits a relative excess of large and small separations.

length≈end-to-end separation) and distinguish them from convoluted random branches (branch length≫end-to-end separation). However, there were too few branches present in these samples to obtain significant results. Analysis of larger data sets might improve the chances of extracting useful information of this sort.

In Fig. 5 we show the frequency distributions of the edge lengths, $F(l)$, in the full MST's for the Zwicky and Poisson samples before pruning and separating. The $F(l)$ distribution for the MST of the random data has a Gaussian form centred on $\langle l \rangle$ while $F(l)$ for the Zwicky sample has an exponential fall-off with a significantly larger contribution at small and large $l/\langle l \rangle$. One of the things this displays is the fact that the Zwicky data is more clustered than the random sample. It also suggests the presence of more subclustering and field objects which result in large l values.

In order to test the stability of these forms for $F(l)$, the full MST was successively pruned down to level 50. At each stage the mean length of all edges remaining on the pruned tree was found and then used to give a new $F(l)$. The forms of the $F(l)$ distributions for both Zwicky and Poisson data sets were found to be remarkably stable at all levels of pruning even though more than 75 per cent of all the nodes were removed during the pruning process.

We have also run several different random realizations of 1091 points, with different values of $\langle l \rangle$ and have found the same form of MST structure illustrated in Fig. 3 and the same Gaussian $F(l)$ form as displayed in Fig. 5. These features appear to be robust and it would be interesting to calculate the expected form of $F(l)$ analytically using the Central Limit Theorem. Such robustness does not exist for percolation statistics (Dekel & West 1985).

Since $F(l)$ just classified the separations between a specially chosen collection of galaxy pairs we would like to know if it contains information that differs significantly from that carried by the two-point angular correlation function, $\omega(\theta)$, (Totsuji & Kihara 1969; Peebles 1980). We note first that, whereas $\omega(\theta)$ measures only average properties of a distribution, $F(l)$ is sensitive to

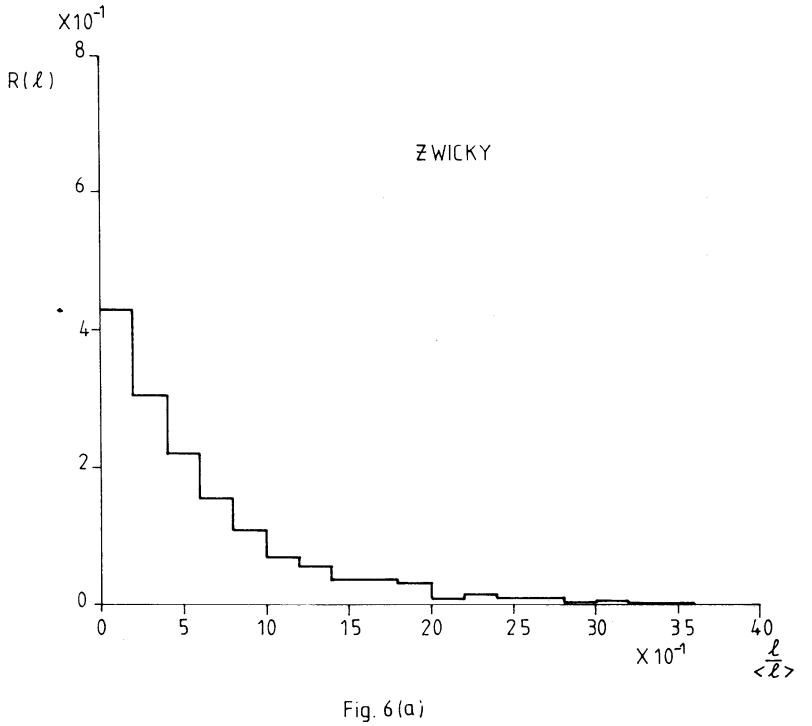

Figure 6. Frequency distribution of two-dimensional MST edge-lengths relative to all two-point separations. The $R(l)$ versus $l/\langle l \rangle$ distributions for the (a) Zwicky and (b) random samples. These distributions normalize the $F(l)$ distributions to allow for the two-point correlations between all points and indicate that the $F(l)$ distributions carry different information about the clustering, especially at small separations.

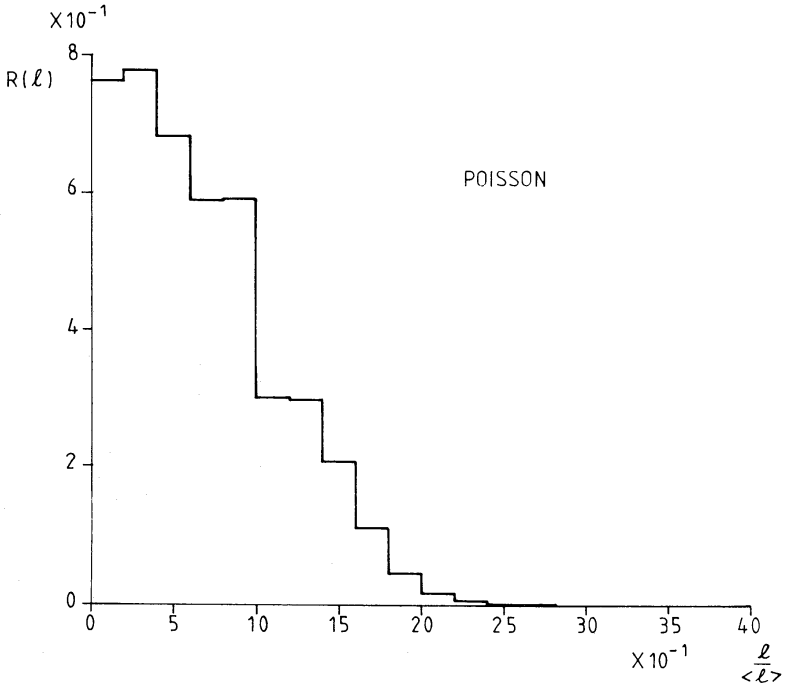

Figure 6 – *continued*

changes in the positions of galaxies and contains a built-in propensity towards describing that distribution of points which dominates the underlying pattern. Although the form $F(l)$ is expected to be influenced by $\omega(\theta)$, the MST selects a special subset of all the possible galaxy pair separations which is strongly biased towards small angles. A useful quantitative measure of the likely difference between $F(l)$ and $\omega(\theta)$ is $R(l)$, *the ratio of the number of separations of length l in the MST of the data set to* N_{00}, *the number of inter-galaxy separations of length l in the whole data set.* The two-point correlation function is determined from N_{00} (see Barrow, Bhavsar & Sonoda 1984, for example). The quantity $R(l)$ can be thought of as a measure of the pattern-connectedness in which the degree of clustering in the underlying data set has been normalized. In Fig. 6 we show $R(l)$ for the random and Zwicky two-dimensional data sets.

4 Application to three-dimensional data sets

The three-dimensional analysis proceeds in the same way as in the previous section except that edge-lengths are now measured as real distances rather than as angles. There is also a practical problem in that we cannot easily display the three-dimensional data. We applied the analysis of Sections 2 and 3 to a volume-limited sample taken from the CfA catalogue (Davis *et al.* 1983), to several Poisson data sets and also to two samples taken from numerical simulations. The subset of the CfA catalogue that we used contained 489 galaxies in the North Galactic Cap, $\delta \geq 0°$, $b^{II} \geq 40°$ with radial distance <80 Mpc; (Hubble velocity $<4000\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$). The positions of these galaxies have been corrected for our Virgocentric infall. The Poisson samples also contained 489 galaxies and occupied the same region as the CfA sample. Our numerical simulation data sets are sub-samples of the 4000 body models of Gott, Turner & Aarseth (1979). The original numerical simulation data sets gave the positions of galaxies inside a sphere. In order that all the results from the different data sets could be compared, the simulation subsamples that we used occupied the same shaped truncated cone as the CfA and Poisson samples. Of the four simulations, we chose

1985MNRAS.216...17B

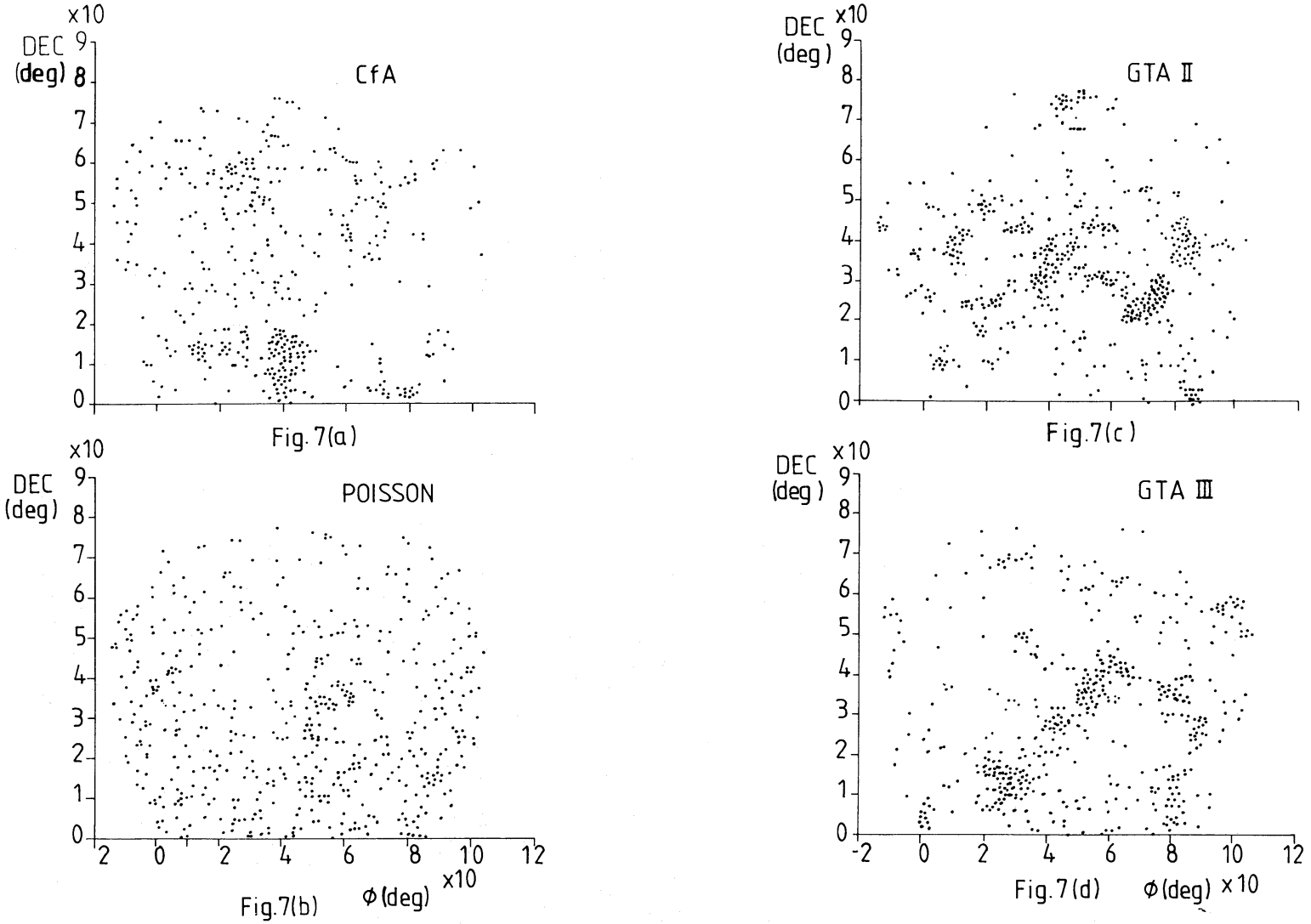

Figure 7. Projections of three-dimensional data sets: (a) CfA Survey, (b) a Poisson sample, (c) the simulation GTA II, and (d) the simulation GTA III. Each three-dimensional data set is displayed as a projection on the celestial sphere. They all occupy the same region of sky, although the sample depths differ. The two simulations used, (c) and (d), possess virtually the same spatial two-point correlation functions as the observations (a), and this is reflected in the visual appearance of the data sets. Again, the mappings of points are not conformal (see the caption to Fig. 3).

models II and III (henceforth denoted by GTA II & III) since their two-point correlation functions, $\xi(r)$, gave the closest match with the observed power-law form $\xi(r)\sim r^{-\gamma}$, $\gamma=1.8$. GTA II and GTA III have different density parameters $\Omega_0=0.1$ and 1 respectively and were evolved from initial perturbations with different power spectra $(\delta\varrho/\varrho)\alpha M^{-1/2-n/6}$; $n=-1$ and $n=0$ respectively. GTA II has $\gamma=1.9$ and GTA III has $\gamma=1.7$, and after truncation contained 1634 and 1320 points. The CfA data set, one of the Poisson samples, GTA II and GTA III are shown as projections on the celestial sphere in Fig. 7. The visual appearance of these point data sets reflects the differences in the value of the two-point correlation function ξ. GTA II has several tight clusters with little filamentary structure; GTA III is more open and does show some linear structure. It should be remembered that the appearance or non-appearance of linear features here may be a consequence solely of the projection.* The three-dimensional MST's were calculated, and the pruned and separated trees are displayed in Figs 8 and 9. The cut-offs used in the separation process are measured in units of the respective mean edge-lengths $\langle l\rangle$ and correspond to the same absolute separation. Although the Poisson 3D MST is noticeably different from the others, the simulation MST's both have the same sort of open structure as the observations. Looking for measures of filamentary structure based solely on the visual appearance of the MST's does not seem too promising since small changes in the cut-offs can have large effects.

The degrees of all nodes, branch lengths and branch end-to-end separations were all examined for the four MST's. As in the two-dimensional case, there were no striking differences in these

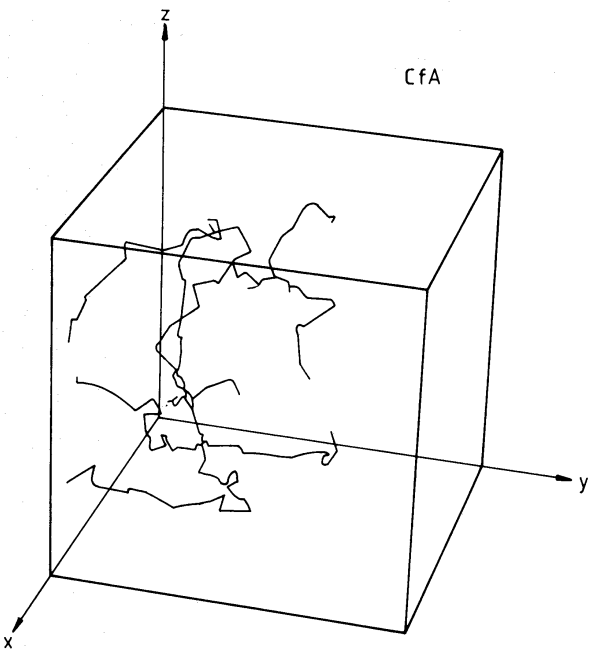

Fig. 8 (a)

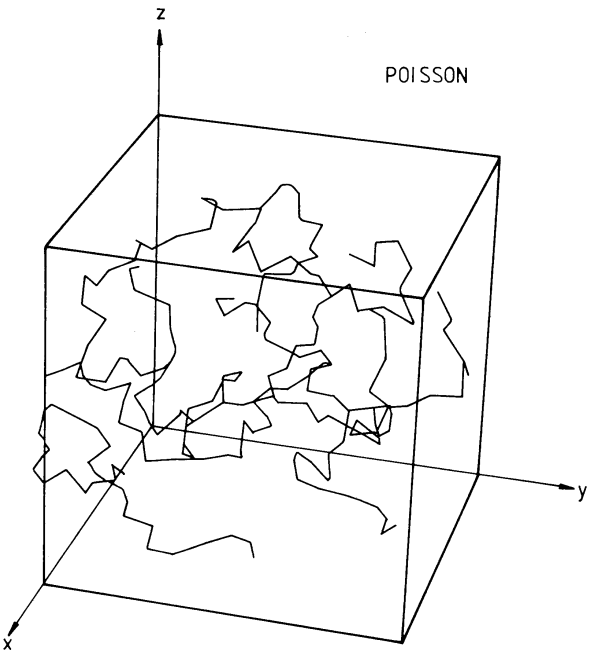

Fig. 8 (b)

Figure 8. Pruned and separated three-dimensional MSTs. (a) the CfA survey, (b) the Poisson sample. For (a) a cut-off of $3\langle l\rangle=10.2$ Mpc was used in the separating process and in (b) a cut-off of $1.8\langle l\rangle$ which corresponds to the same absolute distance cut-off as in (a). In both cases, only those branches with at least 10 edges are displayed in these perspective projections. Although (b) appears far more knotted than (a), this feature is dependent upon the relative sizes of the cut-offs used in each case.

* In order to evaluate the significance of distinctive features in the 3D MST's one should re-evaluate the results from the simulations after calculating galaxy positions using the actual velocity distributions [this was done for the percolation statistics by Bhavsar & Barrow (1984)]. The effects of peculiar velocities can then be evaluated.

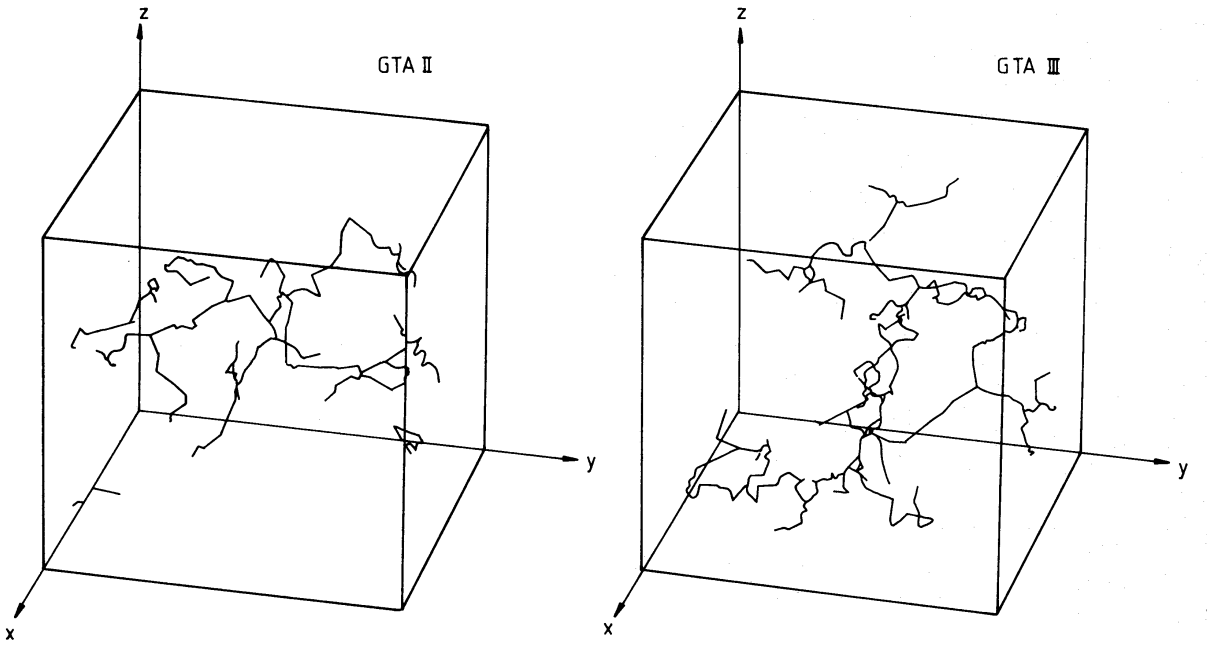

Fig. 9 (a) Fig. 9 (b)

Figure 9. Pruned and separated three-dimensional MSTs: simulations. (a) the simulation GTA II, (b) the simulation GTA III. (a) has a spatial correlation function with slope 1.9 and is therefore slightly more clustered than the observational data whilst (b) is slightly less clustered than the observational data and has a correlation function with slope 1.7. Cut-offs of $3\langle l\rangle_{II}$ and $3\langle l\rangle_{III}$ were used in (a) and (b) to enable direct comparison with the results shown in Fig. 8.

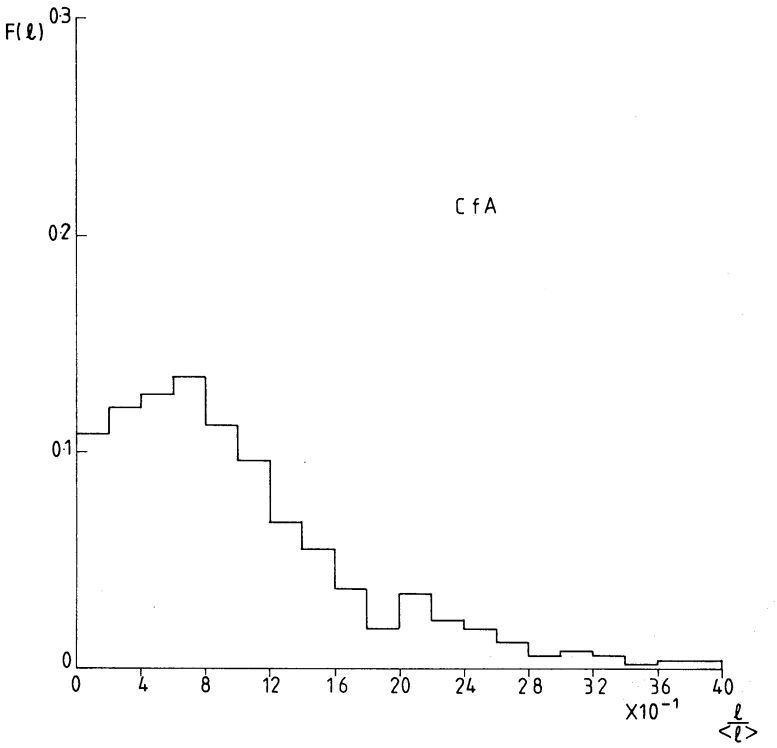

Fig. 10 (a)

Figure 10. Frequency distribution of three-dimensional MST edge-lengths. The $F(l)$ versus $l/\langle l\rangle$ distributions for (a) the CfA survey, (b) the Poisson sample, (c) the simulation GTA II, and (d) the simulation GTA III.

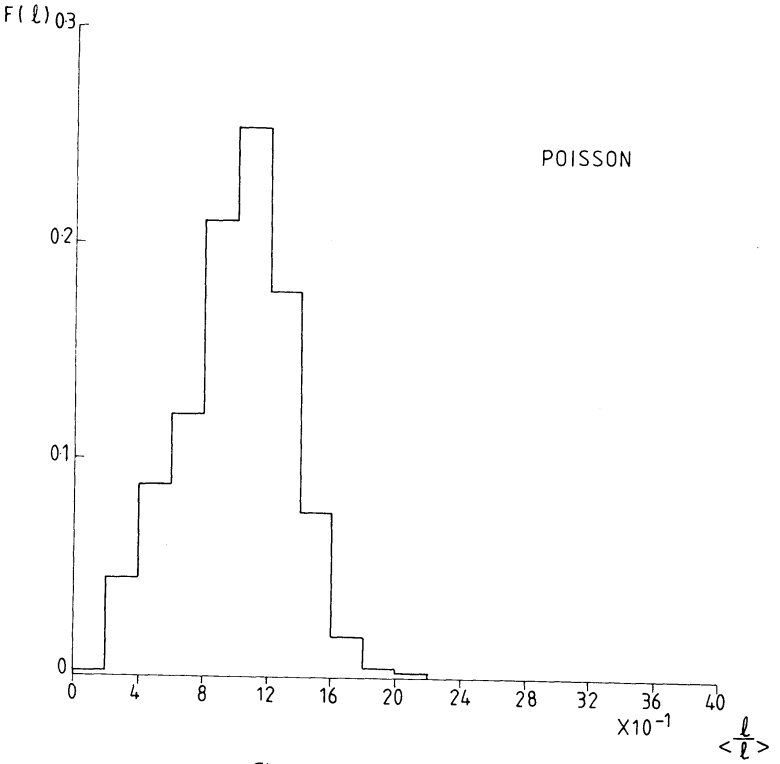

Fig. 10 (b)

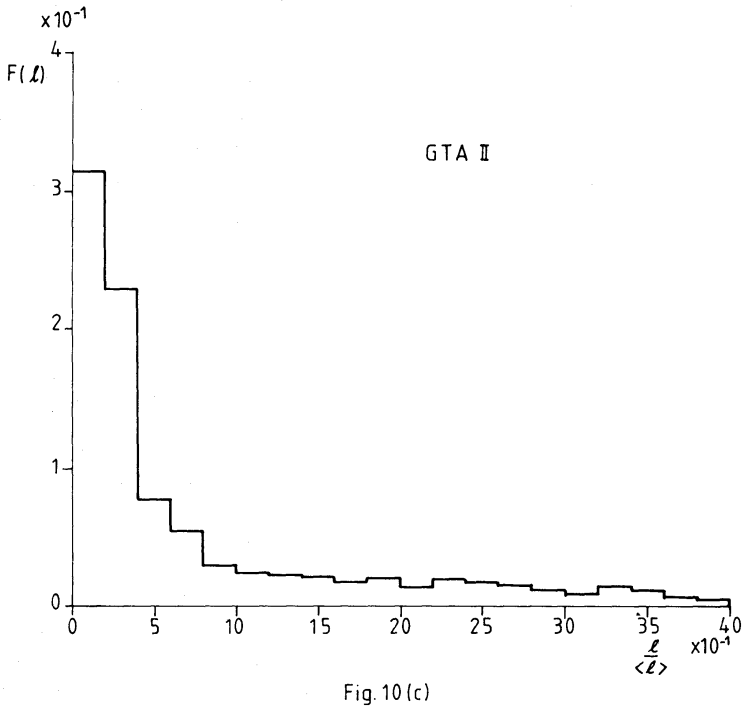

Fig. 10 (c)

Figure 10 – *continued*

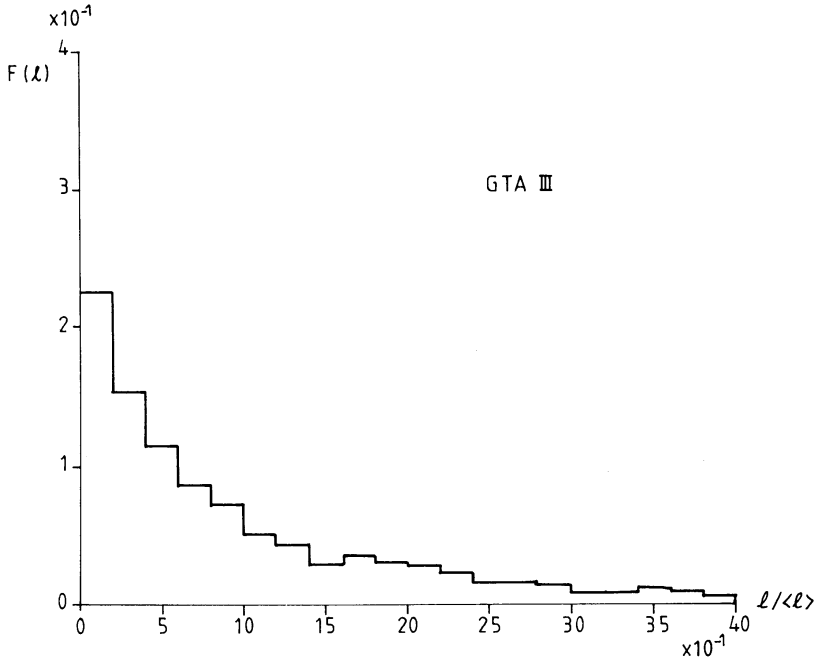

Figure 10 – *continued*

features for the different data sets. We concentrated therefore mainly on the $F(l)$ and $R(l)$ distributions. These are shown in Figs 10 and 11. Just as in the 2D case, the Poisson sample $F(l)$ was well described by a Gaussian distribution centred on $\langle l \rangle$ and the CfA sample had a relative excess of large and small edge lengths. Both simulations had an almost exponential fall-off in their $F(l)$ curves. GTA III decayed more slowly than GTA II and this is, perhaps, expected since GTA II has a steeper correlation function. The $R(l)$ distributions show how the MST preferentially selects small separations.

From $\xi(r)$ alone, GTA II and GTA III give equally good agreement with the observations. However, from the $F(l)$ and $R(l)$ distributions GTA III gives a slightly better agreement with the CfA sample.* There is one feature of the CfA $F(l)$ distribution that neither of the simulations reproduce. Although all the $F(l)$ distributions are well-described by an exponential tail for $l/\langle l \rangle > 1$, there is a noticeable fall-off at small l values in the CfA sample.

5 Conclusions

We have explored some of the consequences of applying a pattern recognition technique drawn from graph theory to the problem of galaxy clustering. A minimal spanning tree (MST) can be constructed systematically and uniquely for any point data set given all the pairwise separations. The resulting MST gives a quantitative measure of the connectedness of a data set that is objective but which gives a close correspondence to the subjective patterns isolated by the eye.† This procedure enables observations and theoretical distributions to be compared. In two-

*It is interesting that the MST analysis favours the $\Omega_0=1$ model, just as our percolation analyses did (Bhavsar & Barrow 1983, 1984). Like percolation analyses, the MST construction picks out different features to the multiplicity function (Bhavsar, Gott & Aarseth 1981). The multiplicity function measures the detailed distribution of the *sizes* (or number) of *all structures* at different density levels, whereas percolation and MST measure *lengths* of the *largest structures* at different levels.

†The fact the filamentary structure is a visual impression is a further motivation for this type of statistical measure.

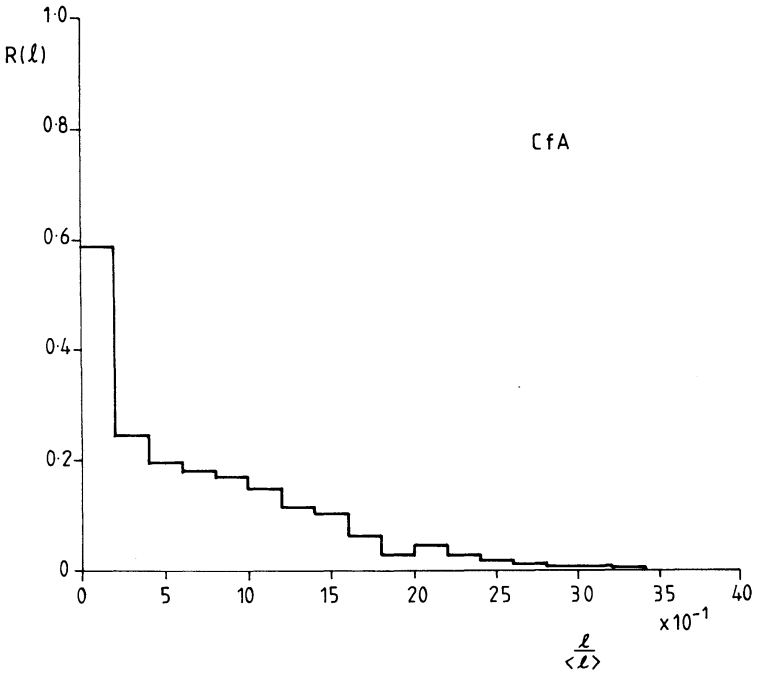

Fig. 11 (a)

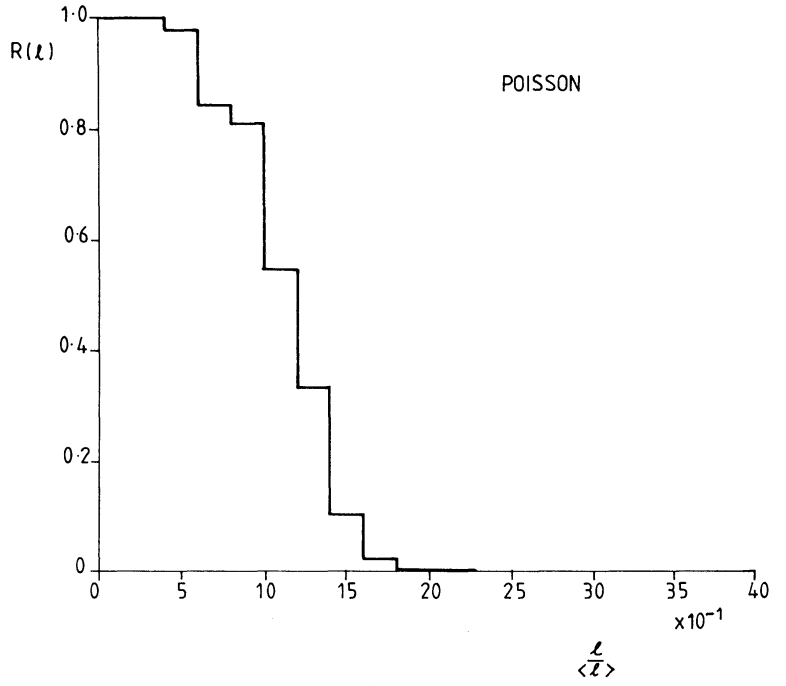

Fig. 11 (b)

Figure 11. Frequency distribution of three-dimensional MST edge-lengths relative to all two-point separations. The $R(l)$ versus $l/\langle l\rangle$ distributions for (a) the CfA survey, (b) the Poisson sample, (c) the simulation GTA II, and (d) the simulation GTA III.

dimensional examples the configuration of the MST is easily visualized and we found that random samples generate characteristic MSTs which differ strikingly from the MST of the Zwicky sample of galaxies. In three dimensions the full MST is not a convenient entity to deal with. However, there exists a variety of properties of the MST which could be studied. In this study we concentrated upon two such properties. We found that the distribution of edge-lengths in the MST of the random points possesses a Gaussian form, centred upon the mean edge-length in both

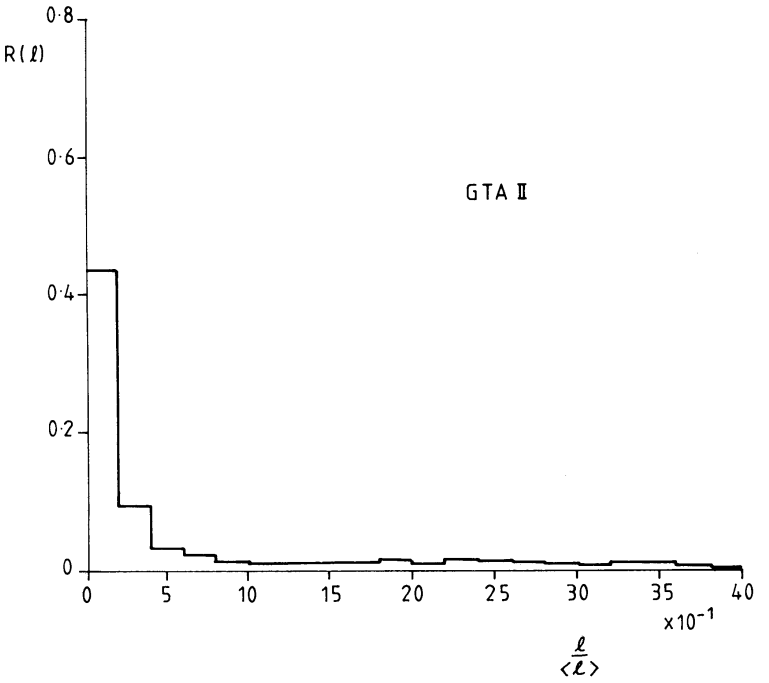

Fig. 11(c)

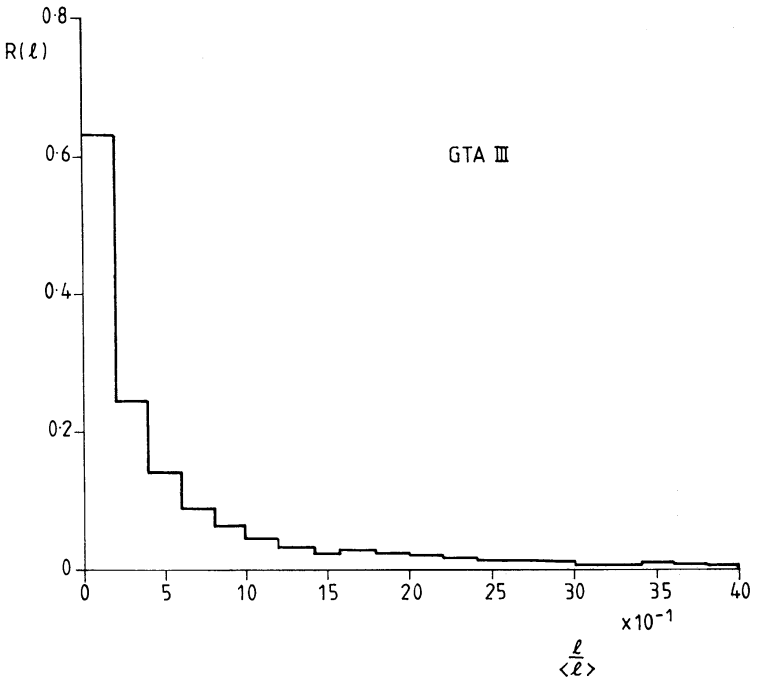

Fig. 11(d)

Figure 11 – *continued*

the two and three-dimensional random samples, but the observations and numerical simulations produced exponential-type distributions with a marked excess of small and very large edge-lengths. In order to evaluate to what extent this distribution function gives different information about clustering than the two-point correlation function we re-evaluated the distribution of MST edge-lengths relative to all two-point separations in the full data set.

The MST is a measure of two-point correlations between a sub-set of the data, specially

selected to trace out the dominant pattern. Although this has only been an introduction to the use of MSTs, the results so far have been encouraging and further study may reveal more sensitive measures of filamentary structure. In a subsequent study we shall report on the application of these methods to the larger *n*-body simulations of Davis *et al.* (1985) containing various possible elementary particles.

Acknowledgments

D. Sonoda was supported by an SERC postgraduate studentship and S. Bhavsar by SERC visiting postdoctoral and senior visiting fellowships at Sussex whilst this work was performed. We are grateful to S. Karbelkar for bringing the UV plot of the Culgoora array (Fig. 1) to our notice and would like to thank C. Frenk, N. Ling and N. Kaiser for discussions.

References

Abraham, C., 1962. *Evaluation of clusters on the basis of random graph theory,* IBM Res. Memo., IBM Corp, York Town Heights, NY.

Barrow, J. D., Bhavsar, S. P. & Sonoda, D. H., 1984. *Mon. Not. R. astr. Soc.,* **210,** 19P.

Bhavsar, S. P. & Barrow, J. D., 1983. *Mon. Not. R. astr. Soc.,* **205,** 61P.

Bhavsar, S. P. & Barrow, J. D., 1984. In: *Clusters and Groups of Galaxies,* eds Mardirossian, M., Giuricin, G. & Mezetti, M., Reidel, Dordrecht, Holland.

Bhavsar, S. P., Gott, J. R. & Aarseth, S., 1981. *Astrophys. J.,* **246,** 656.

Clark, R. & Miller, W. F., 1966. *Math. Comp. Phy.,* **5,** 47.

Davis, M., Huchra, J., Latham, D. W. & Tonry, J., 1983. *Astrophys. J. Suppl.,* **52,** 89.

Davis, M., Efstathiou, G., Frenk, C. & White, S. D. M., 1985. *Astrophys J. Suppl.,* **57,** 241.

Dekel, A. & Aarseth, S. J., 1984. *Astrophys. J.,* **283,** 1.

Dekel, A. & West, M. J., 1985. *Astrophys. J.,* **288,** 411.

Einasto, J., Klypin, A. A., Saar, E. & Shandarin, S. F., 1982. *Mon. Not. R. astr. Soc.,* **206,** 529.

Erdös, P. & Renyi, A., 1960. *MTA Mat. Kut. Int. Közl,* **5,** 17.

Gott, J. R., Turner, E. L. & Aarseth, S. J., 1979. *Astrophys. J.,* **234,** 13.

Gower, J. C. & Ross, G. J. S., 1969. *Appl. Stat.,* **18,** 54.

Iyanaga, S. & Kawada, Y., (eds) 1980. *Encyclopaedic Dictionary of Mathematics,* Vol. I, §69, MIT Press, Cambridge Massachusetts.

Kruskal, J. B., 1956. *Proc. Am. Math. Soc.,* **7,** 48.

Kuhn, J. R. & Uson, J. M., 1982. *Astrophys. J.,* **263,** L47.

Moody, J. E., Turner, E. L. & Gott, J. R., 1983. *Astrophys. J.,* **273,** 16.

Ore, O., 1962. *Am. Math. Soc. Colloq. Publ.,* **38**.

Peebles, P. J. E., 1980. *The Large Scale Structure of the Universe,* Princeton University Press, New Jersey.

Totsuji, H. & Kihara, T., 1969. *Publs astr. Soc. Japan,* **21,** 221.

Zahn, C. T., 1971. *IEEE Trans. Comp.,* **C20,** 68.

Zeldovich, Y. B., Einasto, J. & Shandarin, S. F., 1982. *Nature,* **300,** 407.

Zwicky, F., Herzog, E., Wild, P., Karpowicz, M. & Kowal, C. T., 1961–68. *Catalogue of Galaxies and Clusters of Galaxies,* California Institute of Technology, Pasadena.