CASA

WORKING PAPER SERIES

HIERARCHY IN CITIES AND CITY SYSTEMS

Michael Batty

Hierarchy in Cities and City Systems[†]

Michael Batty

University College London
m.batty@ucl.ac.uk

12 November 2004

Abstract

Hierarchy is implicit in the very term city. Cities grow from hamlets and villages into small towns and thence into larger forms such as 'metropolis', 'megalopolis' and world cities which are 'gigalopolis'. In one sense, all urban agglomerations are referred to generically as cities but this sequence of city size from the smallest identifiable urban units to the largest contains an implicit hierarchy in which there are many more smaller cities than larger ones. This organisation approximately scales in a regular but simple manner, city sizes following a rank-size rule whose explanation is both mysterious and obvious. In this chapter, we begin with a simple but well-known model of urban growth where growth is randomly proportionate to city size and where it is increasingly unlikely that a small city becomes very big. It is easy to show that this process generates a hierarchy which is statistically self-similar, hence fractal but this does not contain any economic interactions that we know must be present in the way cities grow and compete. We thus modify the model adding mild diffusion and then note how these ideas can be fashioned using network models which generate outcomes consistent with these kinds of order and scaling. We then turn this argument on its head and describe how the same sorts of morphology can be explained using ideas from central place theory. These notions are intrinsic to the way cities evolve and we conclude by noting how city design must take account of natural hierarchies which grow organically, rather than being established using top-down, centralized planning.

[†] This paper is to be published in Denise Pumain (Editor) *Hierarchy in Natural and Social Sciences*, Kluwer Academic Publishers, Dordrecht, The Netherlands, forthcoming 2005.

"To a Platonic mind, everything in the world is connected to everything else- and perhaps it is. Everything is connected but some things are more connected then others. The world is a large matrix of interactions in which most of the entries are close to zero, and which by ordering those entries by their orders of magnitude, a distinct hierarchic structure can be discerned." (page 258)

H. A. Simon (1977) *Models of Discovery and Other Topics in the Methods of Science*, D. Reidel Publishing Company, Dordrecht Holland, Boston, MA.

1 The Urban Soup

Conventional wisdom concerning the origins of life on earth are now largely fashioned around the notion that in the beginning, life began through some chance spark setting off a reaction in a sea of undifferentiated chemical soup, leading to the formation of the various nucleotides that constitute the buildings blocks of life – RNA and DNA. In the same way, we can speculate that societies and cities began with household units randomly located across a landscape where the spacing of individuals was determined by food available from hunting and gathering. These units of course made contact in their quest for survival and although the dominant mode was one in which households competed with one another for territory which was synonymous with survival, there was a dawning realisation that cooperation rather than competition could ensure greater prospects for survival. Hamlets and villages were formed initially to ensure strength in numbers, for protection, but in time, the social contact which resulted, reinforced a division of labour leading to increased prosperity.

The simplest possible model is one in which some individuals in this undifferentiated urban soup grow more than others simply due to the fact that they continually get ahead, while others fall behind, often disappearing. Eventually clusters that are differentiated by size which we call cities, appear in this landscape, and it is these that give structure to the urban soup. Hierarchy is an intimate part of this structure but before we show how such hierarchies emerge as a natural part of the growth process, we will take one step back and show how cities in this artificial world first organise themselves according to size.

A hierarchy is a natural ordering that is initially based on size but size can be measured in many different ways. In cities, size is typically based on the number of individuals or households or workers – on populations – but it may also be based on the area over which such location occurs or energy is used, or the field of influence over which individuals in the hierarchy have control. Let us begin with the simplest of possibilities: places of the same size are randomly scattered over a uniform plane. In such a world, if clusters exist, then these are random occurrences. We will assume a process in which a place grows randomly but this growth rate is applied proportionately to the size of that place. So if a place i at time t has size P_{it} and the growth rate ε_{it} is chosen randomly, the place grows (or declines) as

$$P_{it+1} = (1 + \varepsilon_{it})P_{it} = P_{it} + \varepsilon_{it}P_{it} \qquad . \qquad (1)$$

The consequences of this process are surprising at first until one pauses to reflect. In a system of many places, the distribution of growth rates will be uniform at any time t over a range from small to large, which might also be from negative to positive. However the chances of any particular place getting a series of very high growth rates allocated to it one after another, and thus growing very big is increasingly small. Equally the same goes for a place getting increasingly small and of course in this model if a place gets too small, it disappears so there is some asymmetry within the process. It is very easy to work out what happens if we apply this growth process to a small number of objects, with random growth rates chosen from a given range, and then apply these using equation (1) over and over again. An increasingly small number of the objects grow big, most remain small, quite a lot disappear but the crucial issue is 'does the resulting size distribution show any kind of order'. In a sense we have anticipated that it does: there are far fewer bigger objects than smaller but let us take a worked example, which although somewhat artificial, graphically demonstrates the point.

Our example is based on a grid of objects of dimension 21 x 21 giving 441 objects or spaces where the initial populations are uniformly distributed with $P_{i0} = 1, \forall i, t = 0$. The rates of growth ε_{it} are chosen randomly in the range $-0.1 < \varepsilon_{it} < 0.1$. The proportionate growth model in equation (1) quickly sorts out the objects into a size distribution and by time $t = 100$, the frequency distribution shows every sign of being lognormal. In fact, we have run

the model for 1000 iterations which we refer to somewhat euphemistically as 'years' and during this simulation there is much movement between those objects in terms of their relative size. The popular way of organising such frequency distributions is by ordering them hierarchically in terms of size and then plotting them in this order which is against their rank. The so-called rank-size distribution or Zipf plot is in fact the counter-cumulative; that is, the rank is the number of objects above a certain size threshold, so the highest ranked population in the hierarchy at rank $r = 1$ is the population $P_{it}(1)$, the second highest population $P_{it}(r)$ is at rank $r = 2$ and so on down the size distribution to $P_{it}(r)$. This rank-size distribution is plotted for $t = 1000$ in Figure 1, and it is immediately apparent that the signature is that of a lognormal distribution where the plot is visualised as a log transformation of population size against rank.
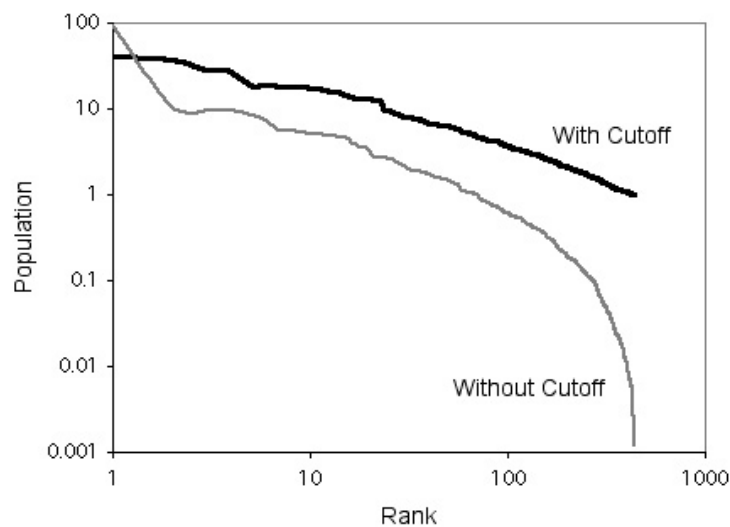

Figure 1: Generating a Lognormal Distribution Using Proportionate Effect and Power Law Scaling from Proportionate Effect with a Minimum Size Threshold

In fact, many researchers have shown that proportionate random growth of the kind we have described leads to lognormal size distributions (Pumain, 2000). Gibrat (1931) produced the first comprehensive argument for cities and income distributions but the English statisticians Fisher and Yule knew of the model and its consequences a generation earlier. If we were to continue the simulation beyond $t = 1000$, then more and more populations would converge to zero and ultimately, we hypothesise that in discrete systems of this kind, all activity would be

attracted to a single cell. In fact, in city systems, such a simulation is bounded from below by indivisibilities and it thus makes sense to modify our model by introducing a size threshold below which populations cannot fall. Then whenever a population cell i falls below this number, it is restored to that number, this mechanism acting as a safety net or subsidy of sorts. This can also be viewed as a way of killing off a city and introducing a new one at the same time, thus incorporating a perfectly balancing birth and death process; formally

$$ if \quad P_{it} < \Psi \quad then \quad P_{it} = \Psi \qquad\qquad . \qquad\qquad\qquad (2) $$

Examining the lognormal distribution in Figure 1 reveals two regimes – the long tail which is almost linear and the short tail which accounts for the order of the smaller settlements. It is tempting to think that the long tail could be approximated by a linear or scaling relation and in effect, if we use the cut-off mechanism just postulated then in a purely phenomenological sense, this is effectively cutting off this short tail. We have run the simulation again with equation (2) now operative and indeed the almost straight line distribution in Figure 1 is that which is generated by time $t = 1000$. The cut-off in fact works and what we end up with is a distribution which is no longer log-normal. In fact it is scaling as it can be approximated by a power function where the population size which we will now call $P_{it}(r)$ varies inversely with the rank r as $P_{it}(r) \sim r^{-\alpha}$ with α the so-called scaling parameter of the distribution. As cities are moving up and down this hierarchy, it is tempting to think of the long tail as a 'steady state' to which cities are 'attracted' and indeed, theorists such as Gabaix (1999) demonstrate that this is indeed the case for the Gibrat process which can converge to the a pure scaling law with the parameter $\alpha = 1$.

2 Rank Size and the Law of Proportionate Effect

What we have just demonstrated has been known as an empirical fact about cities and many other distributions for over one hundred years. The most popular exposition of rank order which conforms to a scaling relation is provided by Zipf (1949) in his remarkable book which examined many such distributions from word frequencies to cities. Zipf argued that these

distributions were not only scaling, conforming to the power law, but also that many such distributions – indeed the implication that all such distributions – were such that the power law was a pure inverse. This means that the rank of the population in continents, countries, and counties, at any level or scale, would conform to $P_{it}(r) \sim r^{-1} = 1/r$. This is the strong form of Zipf's Law. Zipf's Law implies that city size distributions are fractal in that if one examines the relationship at any scale, then the distribution is the same. This is self-similarity in its pure form which in terms of a power law means that if the distribution is rescaled, then this is simply a scaling up of the original distribution. Imagine that the rank size is rescaled by rank to another order s, then the rank size scales as $sP_{it}(r) \sim (sr)^{-1} = s^{-1}r^{-1} \sim r^{-1} \sim P_{it}(r)$ which implies that the scaling is the same over any order of magnitude (Batty and Shiode, 2003).
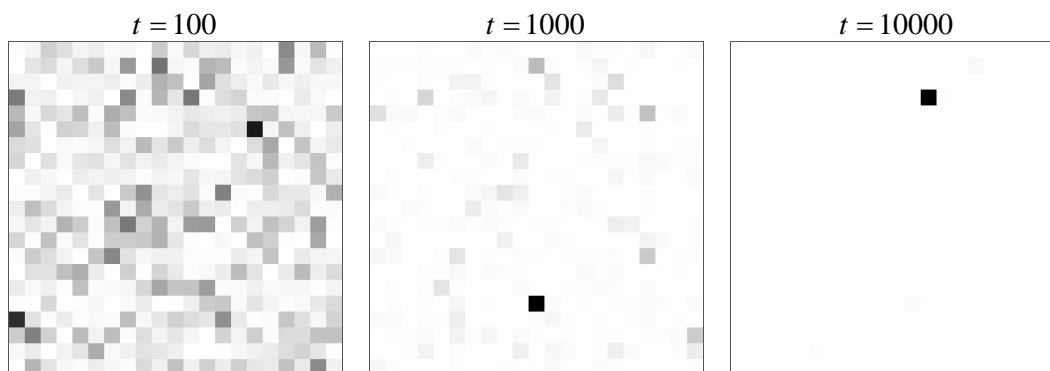
$$t = 100 \qquad t = 1000 \qquad t = 10000$$

Figure 2: Emergence of the Rank-size Distribution Using Proportionate Effect with Cut-Off

The law of proportionate effect with a lower bound is akin to a random walk with a reflecting barrier (Sornette, 2000). The model is simplistic, perhaps nihilistic in that it does not include any form of competition or interaction between the objects. This is extremely odd as cities compete and interact and many models of their formation emphasise such interactions. Our model of proportionate effect with the lower bound clearly generates distributions which appear to be scaling and follow Zipf's Law, but in many ways this model is unstable. The time over which such distributions emerge and the volatility of the top ranked cells or places is sufficient to suggest that the model does not have enough inertia to mirror real places. The fact that it produces size distributions which concur with reality is not sufficient to mean that this is a good model. For example, as we move through the time periods, then the distributions which are generated change not in their scaling but in their shape. We can see

this in two ways. In Figure 2, we show the pattern of distribution after 100, 1000 and then 10000 iterations ('years') which reveals the convergence to extreme distributions as the simulation continues. In Figure 2, the distributions for small number of interactions are much flatter and gentle than those for larger numbers.
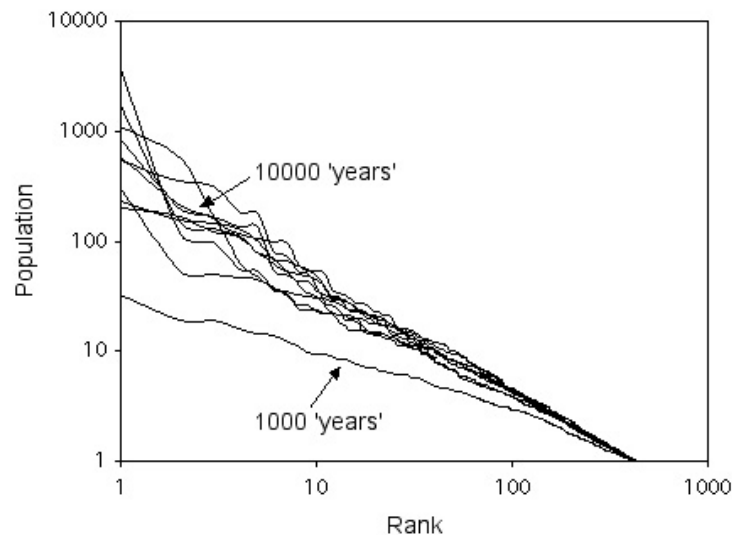

Figure 3: Power Law Scaling as the Population Distribution Emerges

Figure 3 however shows the distributions for $t = 1000, t = 2000, \dots t = 10000$ which clearly get steeper – implying the parameter α gets larger as the population grows. In fact it appears that the parameter is converging on the pure Zipf case of unity although from these results this is inconclusive. There is a variety of theoretical evidence that suggests this is the case for growth by random walk with a reflecting barrier as Gabaix (1999) and Blank and Solomon (2000), for example, show. In Figure 3, the parameter α rises from $t = 1000$ to $t = 10000$ as 0.668, 0.862, 0.907, 0.977, 1.008, 0.984, 0.980, 0.978, 1.053, to 0.962 which shows the final value hovering around 1 with the straight line logarithmic fits all explaining more than 99 percent of the variance for each time slice. This is a fairly remarkable result. What we have shown is that an almost nihilistic model with no spatial competition can generate highly ordered simple hierarchies which in fact mirror the empirical evidence that has been compiled for many cities in many places during the last 50 years. Figure 4 shows the rank size of incorporated places in the United States from 1970 (some 7000 places) to the year 2000 (some

6

25000). In Figure 4(a), the entire distributions are shown and these are clearly lognormal. When we cut-off the short tails, the remaining long tails are quite straight implying power laws as in Figure 4(b). In fact Figure 4 is an empirical equivalent of Figure 1, and for the four time slices from 1970 to 2000, the α parameter varies from 0.986 to 0.982 to 0.995 to 1.014 with the variance explained a little lower than the theoretical model as 0.98, 0.97, 0.97, and 0.97. The same kind of dynamic analysis has been done for France by Guerin-Pace (1995) and a thorough review is presented by Pumain (2005) in a complementary chapter within this book
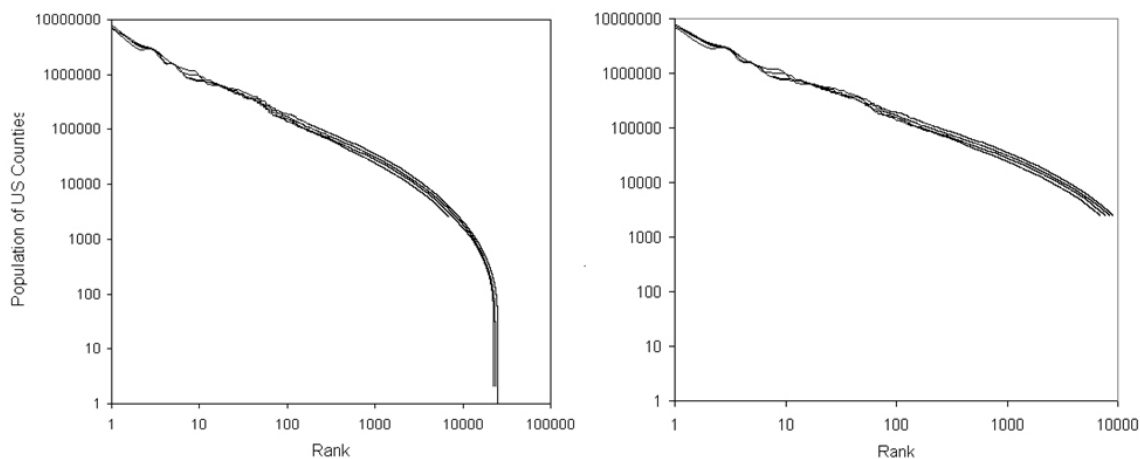

Figure 4: Lognormal and Power Law Scaling of the US Population Based on 'Incorporated Places' 1970 to 2000

Although the model produces aggregate distributions uncannily close to those we observe in most places, what is quite clear is that when we unpack the simulations, there are many inconsistencies that imply this model is nothing like as good as these results suggest. Of particular concern is the lack of apparent structural consistency as the simulation proceeds and the rankings of cells change. During the 10000 'year' simulation, the number of different cells at the top of the rank order is 18. We have only sampled the rankings at every 50 time periods and thus it is likely that there are many more than 18 cells which appear at the top of the ranks during the simulation. To give an idea of the volatility of these ranks, we show how the top ranked cells 1, 6, 12 and 18 from these top orders change over the 10000 year history in Figure 5. These cells appear at different times as we indicate but what is quite clear is that the length of time they occupy the top position is small, thus implying that there is no inertia in

the model. This is manifestly the case and as the simulation time periods have no relationship to real times, then it is unclear what the 10000 model 'years' actually mean in terms of the evolution of actual urban systems such as the US system shown in Figure 4. Our guess is that the volatility of the distribution through time is much greater than any real system where change is slower and inertia greater.
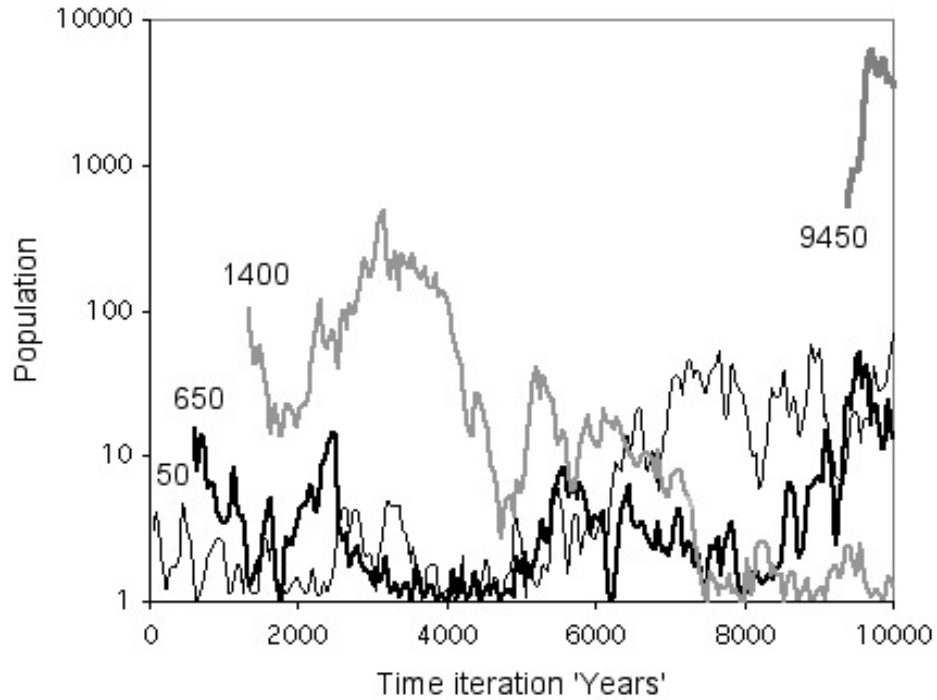

Figure 5: The 1ˢᵗ, 6ᵗʰ, 12ᵗʰ and 18ᵗʰ Top Ranked Population Cells and Their Progress Through the Simulation

The last thing that we want to show before we try to improve the model and generate hierarchies which imply spatial interaction and competition, is the effect of changing the geographic dimensions of the space within which the simulation takes place. We have changed the grid from 21 x 21 to 51 x 51 and then to 101 x 101 and run the simulation with the cut-off for 10000 'years'. We show the three rank-size distributions in Figure 6 where it is quite clear that the slopes are similar implying that the model does indeed hold up as we scale the system in geographic size. This might be expected as there is no interaction between the parts but what is of interest is the increased size of the populations as the systems scales. This is a bit of a mystery but it probably occurs because there are more and more opportunities for

8

extreme growth as the spatial system gets larger, yet it requires further investigation as do many other aspects of these simulations. It does not however detract from the main result that the model scales spatially. It is surprising that so simple a model which has had so much effort devoted to it in terms of simulations and mathematical analysis is still far from being thoroughly understood.
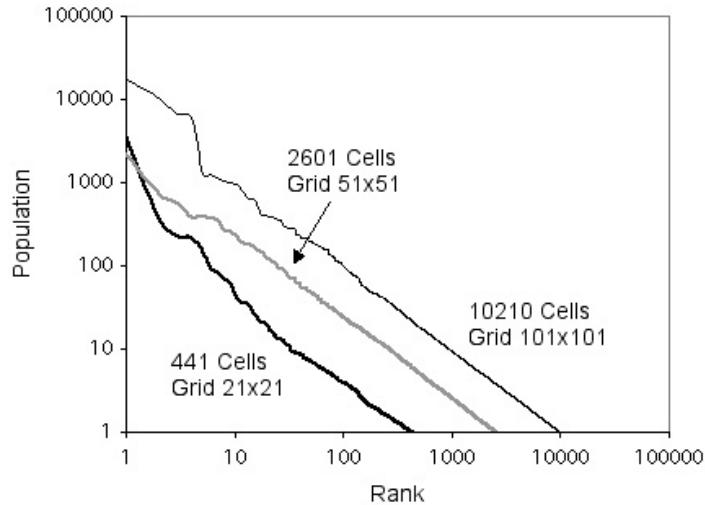

Figure 6: Consistent Scaling Behaviour for Different Sizes of Lattice

3 Generating Hierarchy: Competition, Interaction and Spatial Diffusion

The hierarchy generated by the model of proportionate effect is the simplest possible – a simple rank order or unidirectional hierarchy where the order of objects is simply one of size and where each object is independent of any other. This cannot be a good model for the growth of cities because is does not admit competition or interaction of any kind. Cities are completely disconnected from one another. Simon's opening quote is largely irrelevant to this definition of hierarchy for nowhere in such a model are there clusters of connected activity which provide the kind of connectivity from which hierarchic structure can be derived. What we require is some form of interaction between cities or places, between the points on the lattice and to explore this, we will add some simple diffusion to adjacent grid cells at each stage of the model simulation. In short, at each time step, a fixed proportion λ of the population in each cell k diffuses to its nearest neighbours in the von Neumann

neighbourhood comprising the cells which are north, south, east, and west of the cell in question. Thus for cell i, the population at time $t+1$, P_{it+1}, is now computed as

$$P_{it+1} = (1 + \varepsilon_{it})P_{it} + \lambda \frac{\sum_{k \in \Omega_i} P_{kt}}{4} \qquad , \qquad (3)$$

where the neighbourhood for diffusion is defined as $\Omega_i = k_{north}, k_{south}, k_{east}, k_{west}$. In this model, minimal action at a distance is admitted and within a short time interval which is proportional to one dimension of the lattice, every cell influences every other cell. This kind of diffusion is still somewhat nihilistic in that it is only based on the notion that a proportion of people move to be with their neighbours, without specifying any particular reason, other than the implication that such movement is social and/or economic.
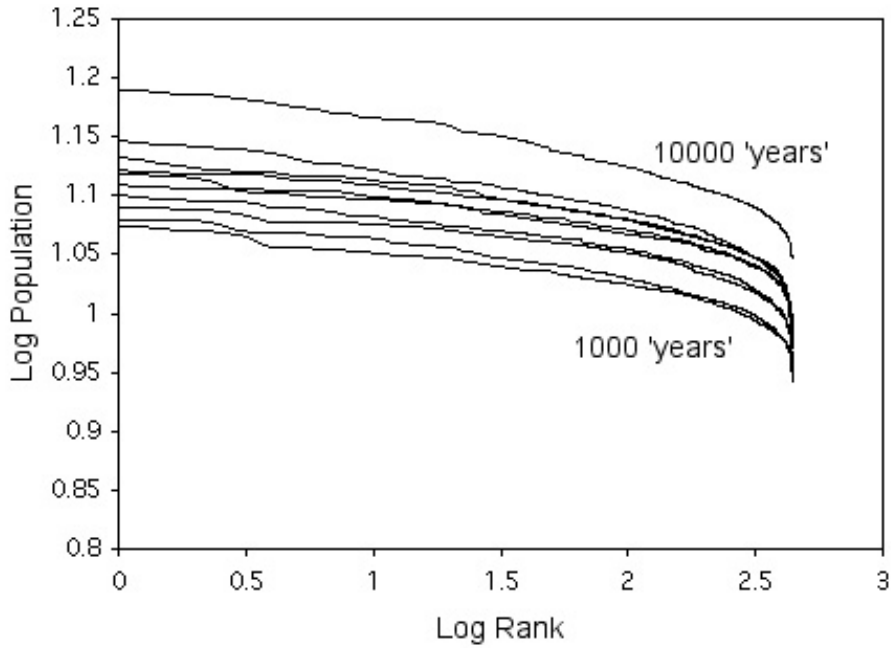

Figure 7: Lognormal Distributions Generated from Proportionate Effect With Diffusion

We have run the model in equation (3) retaining the cut-off in equation (2) for $t = 1000, 2000, ..., 10000$ time periods, and this generates the rank-size distributions shown as

Zipf plots in Figure 7. The level of diffusion used involves setting the parameter $\lambda = 0.3$, implying 30 percent of the population in each cell gets redistributed into adjacent cells in each time period. What this leads to are not scaling distributions as the model without diffusion does but lognormal distributions. The cut-off is in fact discounted by the diffusion. This is a little different from the similar model posed by Manrubia and Zanette (1998) which is based on the same process but without a cut-off and with only positive growth which produces a scaling law. Yet in a sense, whether these kinds of model produce scaling or lognormal distributions is of less concern because at a phenomenological level, all multiplicative processes such as these variants belong to the same class of model (Sornette and Cont, 1997).

The diffusion in this model is so intensive in each time period – all cells are affected, and also extensive due to the fact that the number of time periods is far greater than the size of the system which in this example is based on the 21 x 21 lattice, that it is impossible to track all interactions which accumulate between all pairs of cells. Action-at-distance occurs through the medium of adjacent cells and the number of combinations of diffusion paths is thus enormous. What effectively this diffusion leads to are densities which fall around the cells with the largest populations just as a city core attracts and diffuses activity around it. We take an impressionist view of the hierarchy formed where we simply plot the hierarchy by associating cells with their higher order centre (based on population size), deciding whether or not they are connected simply through adjacency. In Figure 8, we first illustrate the patterns of growth for the model at $t = 100$, $t = 1000$, and $t = 10000$ and it is clear that the type of pattern produced occurs within 100 time periods and simply repeats itself – in different locations of course – through time. In Figure 8, we also show simplifications of these pictures by first identifying the top ranked cell, the next 3 followed by the next 8, then the next 24, and finally the next 64 around these cores. This provides a crude picture of population density which we can represent as a hierarchy. We do this for the pattern at $t = 10000$ where we simply associate each cell at each level with the cells above it if they are connected directly or indirectly to that level through cells of similar value.
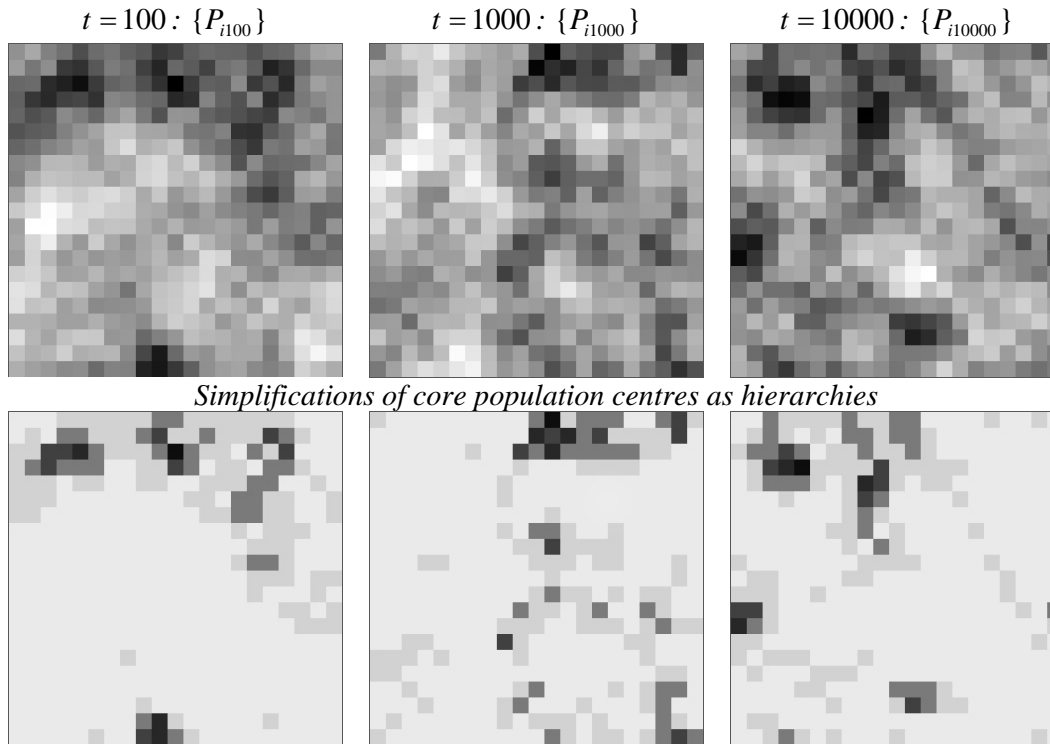
Simplifications of core population centres as hierarchies

Figure 8: Patterns of Diffusion

The hierarchy produced is plotted as a semi-lattice in Figure 9. It is not possible to uniquely associate every cell with a single cell at the next level of hierarchy due to the fact that we do not have network links between cells that we can cut to define separate regions. In fact this representation of hierarchy is much more realistic and supports the long standing notion of overlapping fields of influence which was articulated rather well by Alexander (1966) almost 40 years ago in his article "A City is Not a Tree". There is considerable structure in this hierarchy which is introduced through the diffusion process but this model appears just as volatile as the pure Gibrat process. Over the simulation period of 10000 'years', of the 441 distinct cities or cells, all these cells occur at the top of the hierarchy at some point while the pattern of these top ranked cells would appear quite random. We show this pattern in Figure 10 where it is quite clear that there are no particular clusters of cells or individual cells that predominate over any others. It is quite clear from these simulations that there is too little inertia in the system to mirror our experience of real city systems for it is most unlikely that for the simulation times used here, all cells would at some point dominate. Other models are thus required.
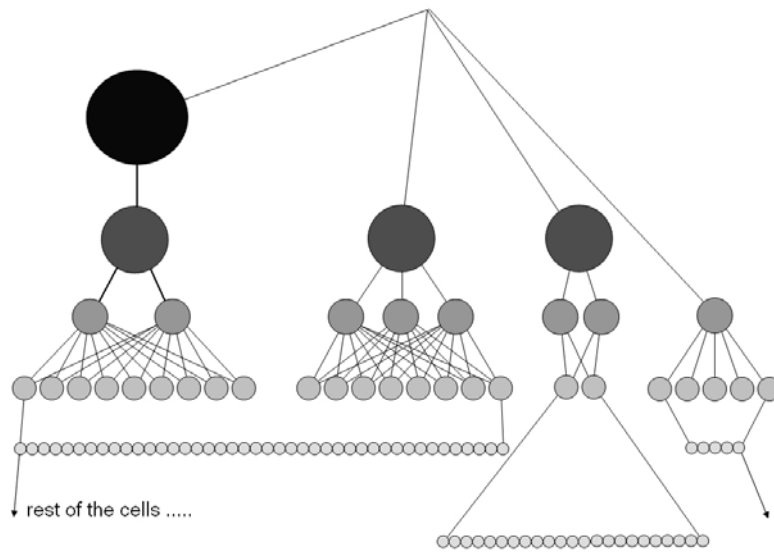
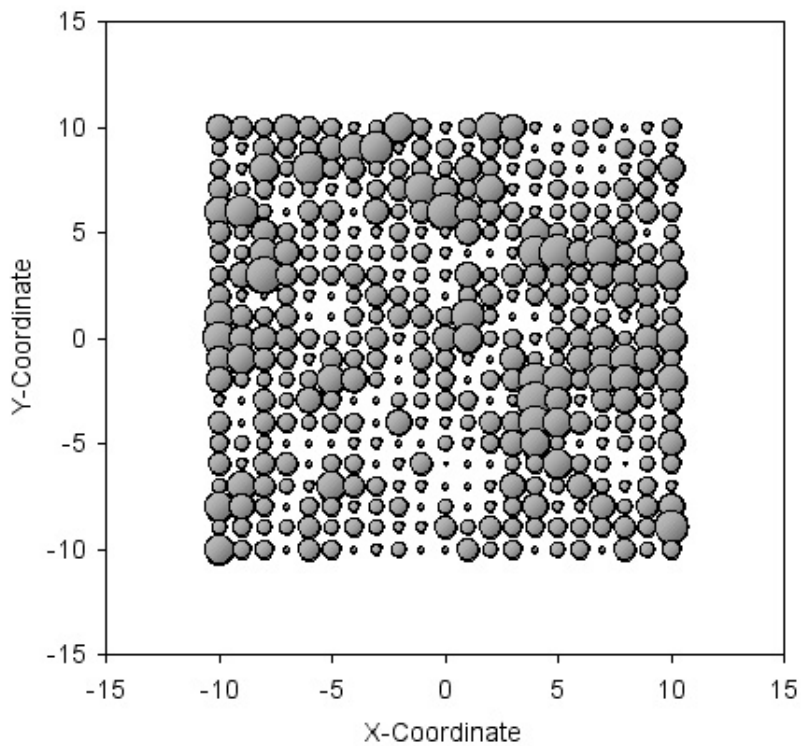*Figure 9: A Hierarchy for the Pattern at $t = 10000$*

Figure 10: Top Ranked Cells During the 10000 'Year' Simulation
(the size of the bubbles range from 1 to 83 time periods in which the relevant cells
dominate with an average size of 23 time periods)

The distributions generated from the Gibrat process with diffusion are somewhat flat and as the level of diffusion is reduced, the hierarchical structure begins to disappear. We have attempted to inject more structure into the model by departing from the Gibrat process and introducing agglomeration economies into the model adding a term reflecting current city size. Our model thus becomes

$$P_{it+1} = (1 + \varepsilon_{it})P_{it} + \lambda \frac{\sum_{k \in \Omega_i} P_{kt}}{4} + \phi P_{it}^{\eta} \,, \tag{4}$$

where ϕ and η reflect the proportionality and the scaling imposed by agglomeration economies. We have set $\phi = 0.2$ and $\eta = 1.08$ and with these parameters we do indeed succeed in sharpening the distribution of city sizes but the lognormality of these distributions remains as we show in Figure 11. There do not appear to be any real qualitative differences produced by this model. To introduce a different form of hierarchy into the urban soup, we require much more explicit networks of interaction exploiting results from the burgeoning science of networks (Watts, 2003) which we will now present.
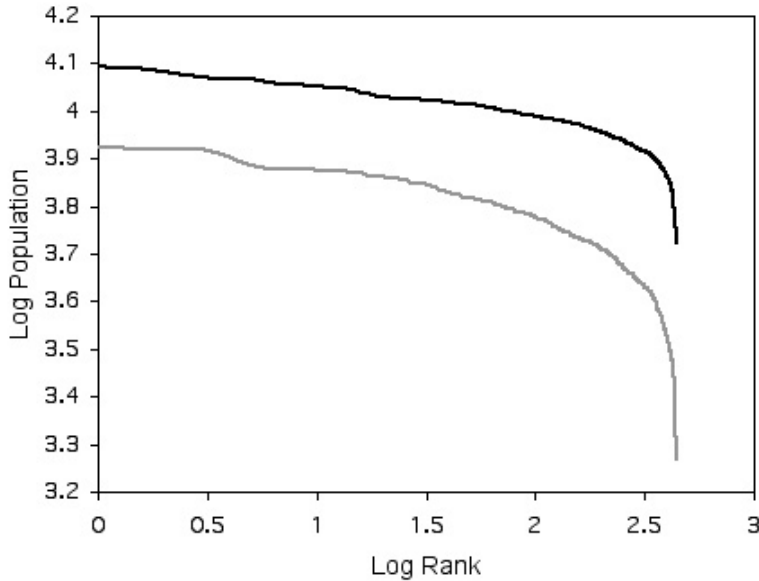

Figure 11: City Size Distributions for the Agglomeration Model at $t = 1000$ and $t = 10000$

4 Network Hierarchies: The Gibrat Interaction Model

So far our models of hierarchy have focused on their evolution whereas Simon's (1977) definition tends to assume that such hierarchies are already developed. To detect them, we thus need to observe the interconnections between the system's parts in order to define the clusters of highly connected subsystems that form the whole. Our model needs to be extended to make explicit these interconnections and this requires us to generalise Gibrat's model to networks. We do this by adding both cells and their links randomly, one in each time period. The mechanics of the model are contained in the following equations. In each time period, for a node which is already established and linked to other nodes, we consider the random addition of a link volume as $\delta_{ikt+1} = 1$ where P_{ikt} is the total number of links from node i to j. The total links associated with i, the new population size of i, is P_{it+1}. The equation for total links is thus

$$P_{it+1} = \sum_j P_{ijt} + \delta_{ikt+1} \qquad , \qquad (4)$$

where the number of links is updated in each time period as

$$P_{ijt+1} = P_{ijt} + \delta_{ijt+1} \qquad . \qquad (5)$$

Whether a link is added or not depends on both the size of the node and its distance to other nodes which is reflected in an exponentially weighted gravitational function of the form

$$\delta_{ijt+1} = \begin{cases} 1 & if \quad rnd(\varepsilon_{ijt+1}) = KP_{ijt} \exp(-\beta d_{ij}) \\ 0 & \end{cases} \qquad . \qquad (6)$$

The term $rnd(\varepsilon_{ijt+1})$ determines a random choice based on size of the potential interaction where d_{ij} is the distance from node i to node j and the parameter β reflects the frictional effects of this distance. Essentially this process is one of preferential attachment in that links

are added in proportion to the size of existing links and the population that the node has already attracted. It has been very widely exploited recently by Barabasi (2002) and his colleagues who have shown that the model does indeed lead to what that call 'scale free' networks (Barabasi and Albert, 1999).

This process does not show how nodes are established in the first place and thus we must add a mechanism for the birth of new nodes, akin to that added by Simon (1955) in his classic model of the rank-size process. A new node is added if a random variable $rnd(\varepsilon_{it+1})$ is greater than a predetermined threshold z which is given as

$$\delta_{it+1} = \begin{cases} 1 & if \quad rnd(\varepsilon_{it+1}) > z \\ 0 \end{cases} \qquad , \qquad (7)$$

where the value of z is small compared to the probability for the addition of new links as reflected in equations (4) to (6) above. For the 21 x 21 lattice, we choose the threshold for the addition of new nodes as $z = 0.1$ and this implies that at the beginning of the process, there is a 1 in 10 chance that a new node is added. Of course as the process continues, this chance falls for if a node is chosen that is already established, this is abandoned. In terms of the generation of links to established nodes, then a node is first chosen randomly but in proportion to its size P_{it}, and then a link to another node j from i is chosen in proportion to its inverse distance function as defined in equation (6). In this way, the network builds up through preferential attachment to existing nodes. The overall dimension of the system is 300 x 300 *x-y* coordinate units for each grid square and thus we have set the deterrence parameter β in equation (6) as 0.001 which implies an average distance of around 1000 units.

In Figure 12, we illustrate the final distribution of population by node $\{P_{i1000}\}$ and alongside this, the distribution of link volumes between nodes for all links greater than 1, those greater than 2 and finally those greater that 4. A hierarchical pattern is revealed by these figures and it would be possible to cut the link volumes at points where the cluster density falls below various thresholds, thus uniquely partitioning the space into different areas and then orders of hierarchy. We do not do this for our concern is not hierarchy *per se* but ways of generating

this. In Figure 13, we plot the size of population per node against their rank as a Zipf plot which is the logarithmic transform. We have not connected the points in this plot because of the comparative low volume levels generated in this example but the plot is mildly lognormal. Fitting a straight line to this gives a scaling parameter of 1.05 with 90 percent of the variation in this plot explained. This is remarkably close to the pure Zipf scaling where $\alpha = 1$ and it is confirmation that this model of preferential attachment based on Gibrat (1931) does indeed generate the same profile as in the simpler non-network cases which we discussed above.
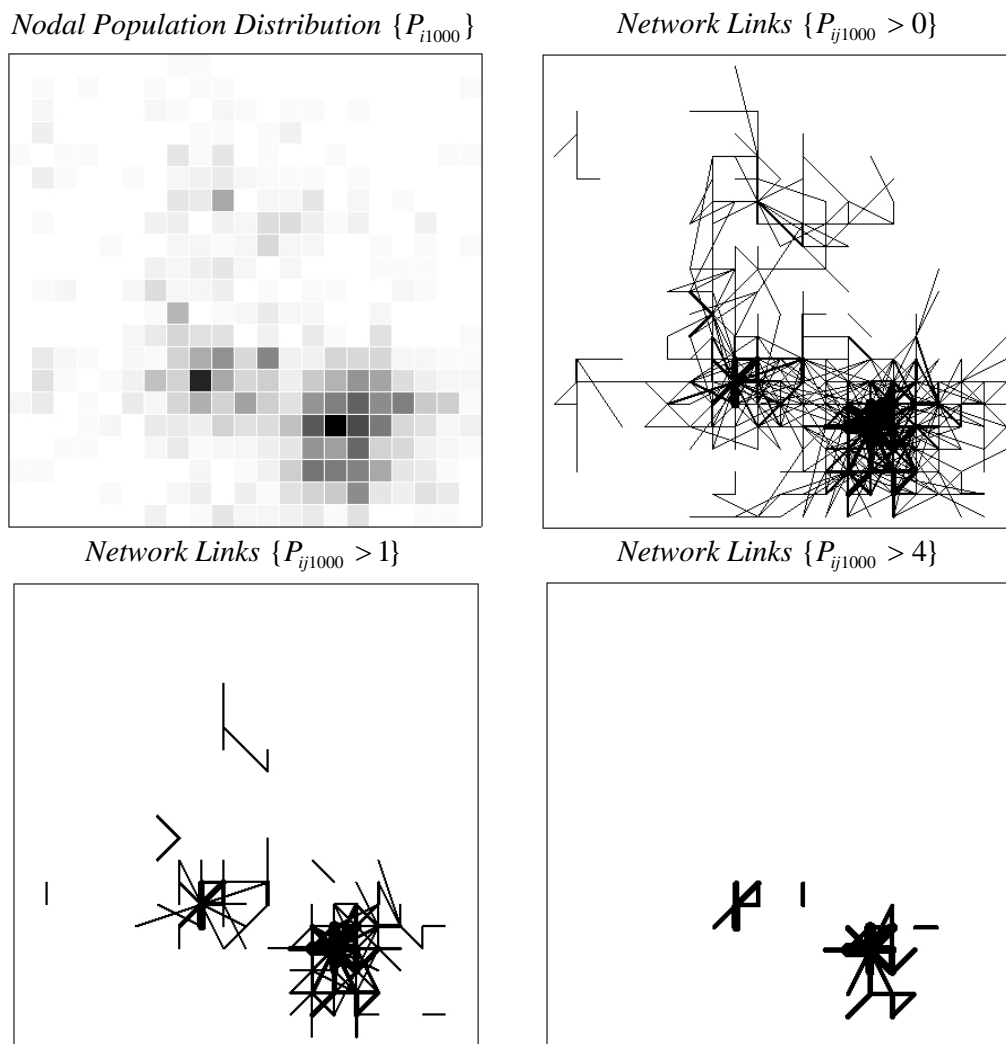
Nodal Population Distribution $\{P_{i1000}\}$ *Network Links* $\{P_{ij1000} > 0\}$

Network Links $\{P_{ij1000} > 1\}$ *Network Links* $\{P_{ij1000} > 4\}$

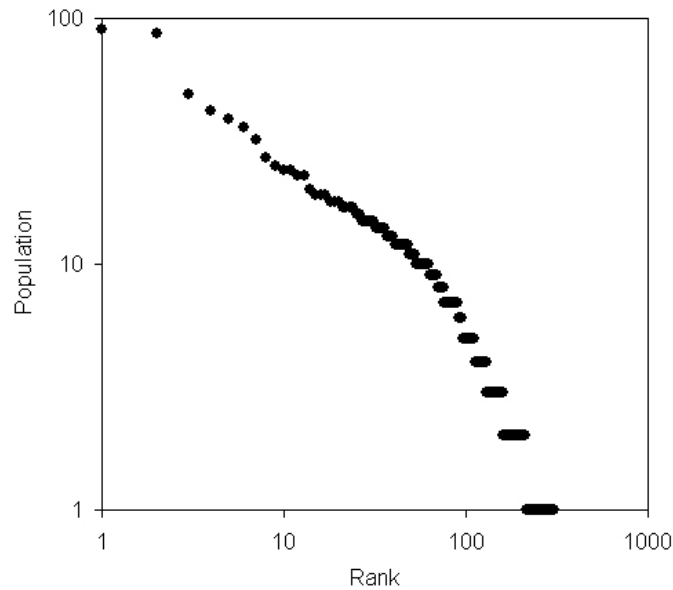*Figure 12: Patterns of Network Connectivity*

Figure 13: Rank-size Distribution of the Nodal Network Distribution

Before we introduce our final model where we will turn back to theory and generate the rank-size rule for hierarchies based on central place theory, we will illustrate an example of hierarchy for retailing within London. We have developed an index of retail intensity which is a linear weighted sum of some 42 separate indicators, each suitably normalised, and tagged to the postcode geography which at its finest scale represents retailing at an average resolution of some 50 metres (Thurstain Goodwin and Batty, 2002). We have interpolated a surface from this data and have then sliced it at some 5 different levels which provides a picture of the retail hierarchy which we show in Figure 14. This is an implicit hierarchy similar to those which can be derived from the population distributions illustrated earlier in Figures 8 and 12. What we do not have from this analysis is the detailed interaction pattern that links consumers to the retailing activity through their movements to purchase retail goods at different points or centres on this surface. But the pattern is consistent with all that we have seen previously and the distribution of retailing activity is rank size. This is only one way of implying that a hierarchy exists and in a sense, this is less explicit than the more formal approaches rooted in location theory to which we will now turn.
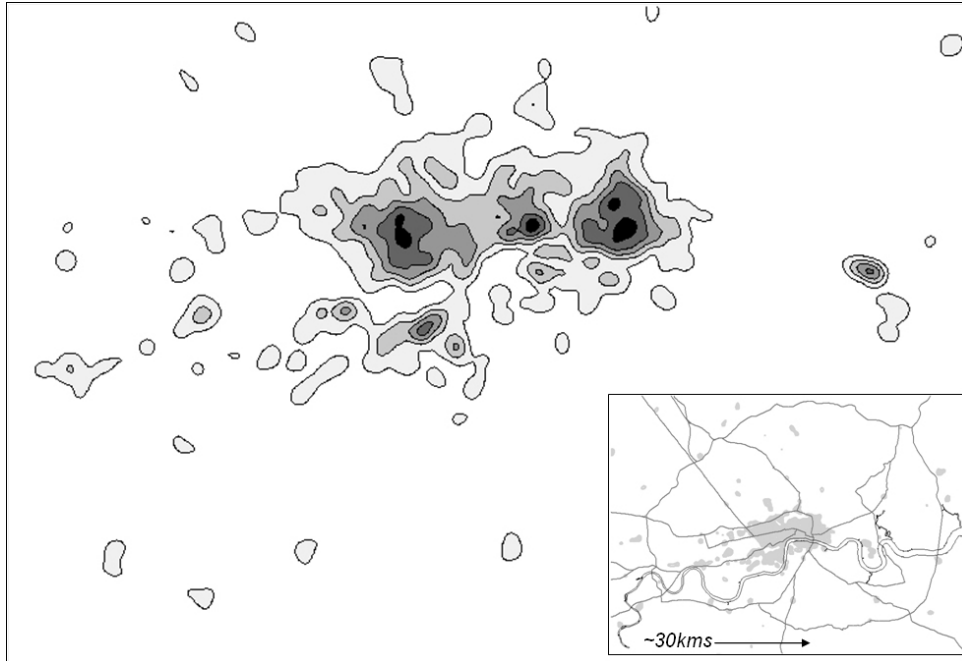
Figure 14: The Implicit Retail Hierarchy in Central London

5 Central Places: Rank Size from Geographical Dependence

Although we have illustrated models which produce quite distinct hierarchies, we have only introduced space as action-at-a-distance from distinct nodes. In so far as competition has entered the argument, this has been either through intersecting and overlapping diffusion or through an implicit ordering where larger places get preferential treatment relative to smaller places, as in the network model of the previous section. One of the first expositions of how geographical areas based on spheres of influence around towns and cities are consistent with the rank-size rule was developed by Beckmann (1958) and his argument is so clear that we will repeat it here, thus providing some sense of closure on our more general discussion of hierarchy through rank-size scaling. Beckmann (1958) defined two key elements in the way cities are organised with respect to their functional and spatial dependence. He first assumed that a city or rather a small seed which sparked off the growth of a city was proportional in size to the population on which it depended in its surrounding hinterland or sphere of influence. He then noted that each city had a 'span of control', which related to the number of lower order hinterlands which could be said to depend spatially and economically on the

19

centre city or seed at its core. This second kind of dependence leads directly to a series of hinterlands at different orders, increasing in number and decreasing in geographical area as they descend the hierarchy and it is from this that the rank of any city can be established.

Formally, the initial dependence ξ of the city seed p_n on its wider population P_n for any order of city n is

$$p_n = \xi \, P_n \qquad , \qquad (8)$$

where the order n is from the largest city which we call N to the smallest which is defined by the index 1. The second spatial dependence involves the fact that the population of the higher order level P_n is a sum of populations at the next lower order P_{n-1} defined as

$$
\begin{aligned}
P_n &= p_n + s\,P_{n-1} \\
&= \frac{s}{1-\xi} P_{n-1} \qquad .
\end{aligned}
\qquad (9)
$$

Recurrence on equation (9) leads to

$$P_n = \left(\frac{s}{1-\xi}\right)^m P_{n-m} \qquad , \qquad (10)$$

and at the bottom of the hierarchy where the population is at the lowest level $P = P_1$, then the exponential dependence within the hierarchy is clear

$$P_n = \left(\frac{s}{1-\xi}\right)^{n-1} P \qquad . \qquad (11)$$

If we assume that the seed city is small or even zero, then $\xi = 0$, and equation (11) simplifies to $P_n = s^{n-1} P$.

Using the reverse order which is from 1 to N, the total number of cities at each level is s^m and the total number up to m is given as

$$\sigma(m) = 1 + s + s^2 + s^3 + \ ... \ + s^m \quad , \tag{12}$$

where this is a diverging but geometric series whose sum is $(s^m - 1)/(s-1)$. Thus the rank of the first city at level m is $\{[(s^m - 1)/(s-1)] + 1\}$ and the rank of the city which is midway through this order – the average rank for this order – is

$$r(m) = \frac{s^m - 1}{s-1} + \frac{s^m}{2} \quad . \tag{13}$$

Examining components of this sum in equation (13), we can assume that $1/s-1$ is small relative to other terms and thus equation (13) can be simplified to

$$r(m) = s^m \left(\frac{1}{s-1} + \frac{1}{2} \right) . \tag{14}$$

The rank-size relation is based on population size and rank and if we multiply the relevant equation for population in equation (11) (which we convert from order n to m as $n = N - m + 1$) with the rank in equation (14), we get

$$P_{N-m+1} \ r(m) = \left(\frac{s}{1-\xi} \right)^{N-m} P \ s^m \left(\frac{1}{s-1} + \frac{1}{2} \right)$$

$$= \Phi \ (1-\xi)^m \tag{15}$$

This equation is a constant if $\xi = 0$ and thus the entire argument hinges on this. To an extent this is arbitrary although it is easy to assume that the hinterland population dominates and the core or seed is near zero. If this holds, we can simplify equation (15) as $P_{N-m+1} \ r(m) = \Phi$. Writing this in a more familiar way where we suppress the order indices and define population at a rank r as P_r, we get $P_r \ r = \Phi$ or

$$P_r \sim \frac{1}{r} \quad , \tag{16}$$

which is the pure Zipf case. Many assumptions have been made to get this far and of course we have not tried to generalise Beckmann's (1958) discrete case to a continuous one. Nevertheless, it would appear that this kind of geographic, indeed geometric reasoning which assumes that space is nested hierarchically through its economic dependence does lead to rank-size distributions of activities such as population. At one level of course, this is all too obvious in that we have assumed hierarchical order and simply shown that geometric series which can describe such order can be manipulated to produce a rank size. This is in fact an indirect argument reflecting scaling. Surprisingly the Beckmann model has not been widely exploited and no one (as far as I know) has developed a stochastic version of it. Nevertheless, it does serve to remind us that there is a deep underlying rationale for the existence of rank-size distributions which is essentially a spatial or geometric ordering in the geographical sense (Beckmann, 1968).

6 Hierarchy in the Design of Cities

The models we have used in this chapter to generate spatial hierarchies whose signature is the scaling of population are essentially stochastic and dynamic although the Beckmann (1958) model of the last section took a more deductive approach but whose dynamics was implicit at best. Yet there are other ways of generating spatial hierarchies. It is possible, for example, to generate such distributions as the outcome of various optimisation procedures, taking either a top-down static approach or even a quasi-dynamic one. By way of conclusion, and in our quest to square the circle and show how hierarchical systems in cities should feature in their design, it is worth noting that there is a long tradition in spatial interaction modelling in which scaling distributions of population and trip/traffic distribution can be derived using optimisation theory – from maximising utility-like or entropy/accessibility functions subject to constraints on the dispersion of such activities through their cost structures. Berry (1964) was one of the first to illustrate such an approach in his derivation of population distributions which conformed to rank size using entropy maximising techniques; and this approach was widely used by Wilson, Coelho, Macgill and Williams (1981) in their quest to embed behavioural land use-transportation models into contexts in which behaviour was considered as optimising either at the individual or collective plan-making level.

In terms of these different approaches, the rank-size distribution provides a sharp illustration of the problems we face in explaining the evolution of complex systems. Nothing could be more different than the generation of a distribution from a stochastic process where all the constituent elements are independent from one another and where the growth is from the bottom up – our first model – and the kind of top down optimisation process in which accessibility is maximised subject to some constraints on cost or energy expended. Yet the outcomes in terms of the distribution of the elements being optimised are the same. In a way, all this shows is that how we approach city systems conditions the techniques we use to generate the outcomes we expect. In terms of the design of cities which although linked to optimisation, originates from very different intellectual mindsets and professional concerns, approaches using ideas from hierarchy theory are also well established. Although many of these use hierarchy in terms of the structure of the problem-solving process where problems are partitioned into a hierarchy of sub-problems, the notion that we need equivalent simplifications to those we have sought here is instructive.

We anticipated much of this in an earlier section where we quoted Alexander (1966) who argued that the notion of strict hierarchy was far too simplistic an organising concept for design. He amongst many others drawing on ideas from organically evolving systems which latterly have been exploited in neo Darwinism by writers such as Dennett and Dawkins, argued for a paradigm in which interaction rather than hierarchy was a required design construct. Overlapping hierarchies – semi-lattices as we illustrated in Figure 9, are much more appropriate vehicles for the organisation of cities into spaces at different levels of geographical scale. In essence, this argument suggests that strict hierarchical subdivision is too simplistic a concept for the design of neighbourhoods and town spaces although it has been widely used by architects operating in top down fashion. Overlapping hierarchies although simplifying interaction capture, the diversity of behaviour and are much more suitable pictorial vehicles for progressing good urban design. In a sense, this argument has also been anticipated in urban systems science; Christaller's central place hierarchies were overlapping while the whole point of spatial interaction modelling and its link to retail centre definition has been to relax the notion of hierarchy, letting it remain implicit in the space of flows.

There are other more direct reasons for thinking of cities as overlapping hierarchies or lattices and this simply emerges from the fact that there are many such hierarchies. We have only examined the simplest here – that based on how the population in an aggregate sense arranges itself but once one disaggregates populations into the multiplicity of categories that define them, and once one adds other kinds of activities which arrange themselves hierarchically in space such as transport and other network systems, land uses, styles of buildings, social friendship nets, and so on, then the idea of overlap becomes the rule not the exception. In fact it is hard to escape from the fact that the best analysis should tackle this notion directly. The fact that most of our analysis tends to simplify the system beyond this obvious reality poses a dilemma. What we require are good, simple and plausible models that show us how different kinds of hierarchies interlock.

In terms of city size distributions, then the challenge seems to be to build on the network characterisations of Gibrat's model, possibly interlocking the network model we illustrated earlier in this chapter with some sort of dual but countervailing network based on friendship patterns rather than the economics of travel which were implicit in the model demonstrated. Interlocking networks which lead to interlocking but consistent and simple scaling of aggregate activities would seem to be the quest. We know that most distributions that we see in cities are scaling or near scaling, and the goal would be to show how these might be unpacked and linked at the network level where we are able to grapple with the diversity that characterises cities. In this way, our understanding of cities would be enriched and this would suggest ways in which we might be able to design our patterns of interaction and location more effectively.

7 References

Alexander, C. (1966) A City is not a Tree, *Architectural Forum*, *122*, (1), 58-61, and *122*, (2), 58-62.

Barabasi, A. (2002) *Linked: The New Science of Networks*, Perseus Publishing, New York.

Barabási, A., and Albert, R. (1999) Emergence of Scaling in Random Networks, *Science*, *286*, 509-512.

Batty, M. and Shiode, N. (2003) Population Growth Dynamics in Cities, Countries and Communication Systems, in P. A. Longley and M. Batty (Editors) *Advanced Spatial Analysis,* ESRI Press, Redlands, CA, 327 –343.

Beckmann, M. J. (1958) City Hierarchies and the Distribution of City Size, *Economic Development and Cultural Change*, 6, 243-248.

Beckmann, M. (1968) *Location Theory*, Random House, New York.

Berry, B. J. L. (1964) Cities as Systems within Systems of Cities, *Papers and Proceedings of the Regional Science Association*, *13*, 147-163.

Blank, A. and Solomon, S. (2000) Power Laws in Cities Population, Financial Markets and Internet Sites: Scaling and Systems with a Variable Number of Components, *Physica A*, *287*, 279-288.

Gabaix, X. (1999) Zipf's Law for Cities: An Explanation, *Quarterly Journal of Economics*, *114*, 739-767.

Gibrat, R. (1931) *Les Inégalités Économiques*, Sirey, Paris.

Guerin-Pace, F. (1995) Rank-size Distribution and the Process of Urban Growth, *Urban Studies*, *32*, 551-562.

Manrubia, S. C., and Zanette, D. H. (1998) Intermittency Model for Urban Development, *Physical Review E*, *58*, 295-302.

Pumain, D. (2005) Alternative Explanations of Hierarchical Differentiation in Urban Systems, in D. Pumain (Editor) *Hierarchy in Natural and Social Sciences*, Kluwer Academic Publishers, Dordrecht, The Netherlands, forthcoming.

Pumain, D. (2000) Settlement Systems in the Evolution, *Geografiksa Annaler*, *82B*, 73-87.

Simon, H. A. (1955) On a Class of Skew Distribution Functions, *Biometrika*, *42*, 425-440.

Simon, H. A. (1977) *Models of Discovery and Other Topics in the Methods of Science*, D. Reidel Publishing Company, Dordrecht Holland, Boston, MA.

Sornette, D. (2000) *Critical Phenomena In Natural Sciences: Chaos, Fractals, Self-Organization and Disorder*, Springer-Verlag, Berlin.

Sornette, D., and Cont, R. (1997) Convergent Multiplicative Processes Repelled from Zero: Power Laws and Truncated Power Laws, *Journal de Physique I (France)*, *7*, 431-444.

Thurstain-Goodwin, M. and Batty, M. (2001) The Sustainable Town Centre, in A. Layard, S. Davoudi, and S. Batty (Editors) *Planning for a Sustainable Future*, Routledge/Spon, London, 253-268.

Watts, D. J. (2003) *Six Degrees: The Science of a Connected Age*, W. W. Norton and Company, New York.

Wilson, A.G., Coelho, J. D., Macgill. S. M., and Williams, H. C. W. L. (1981) *Optimisation in Locational and Transport Analysis*, John Wiley and Sons Ltd, Chichester, UK.

Zipf, G. K. (1949) *Human Behavior and The Principle of Least Effort*, Addison-Wesley, Cambridge, MA.