CENTRE FOR ADVANCED SPATIAL ANALYSIS

Working Paper Series

Paper 64

THE EMERGENCE OF CITIES: COMPLEXITY AND URBAN DYNAMICS

Michael Batty

UCL

Centre for Advanced Spatial Analysis
University College London
1-19 Torrington Place
Gower Street
London WC1E 6BT

[t] +44 (0) 20 7679 1782
[f] +44 (0) 20 7813 2843
[e] casa@ucl.ac.uk
[w] www.casa.ucl.ac.uk

http//www.casa.ucl.ac.uk/working_papers/paper64.pdf

Date: July 2003

ISSN: 1467-1298

The Emergence of Cities: Complexity and Urban Dynamics

Michael Batty

Centre for Advanced Spatial Analysis, University College London,
1-19 Torrington Place, London WC1E 6BT, UK

m.batty@ucl.ac.uk

THE EMERGENCE OF CITIES: COMPLEXITY AND URBAN DYNAMICS

Michael BATTY
Professor and Director
Centre for Advanced Spatial Analysis
University College London
1-19 Torrington Place
London WC1E 6BT, UK
Tel: +44-207-679-1781
Fax: +44-207-813-2843
E-mail: m.batty@ucl.ac.uk

May 29, 2003

Abstract: This paper presents an approach to urban dynamics that generalizes the traditional rank-size model first popularized by Zipf (1949). It argues that we need to define the rate at which new cities emerge and old cities disappear within the apparent macro stability posed by Zipf's Law. We illustrate this with a reworking and extension of Zipf's analysis of the US urban system, taking his analysis from 1790 to 1930 forward to the year 2000. In doing so, we introduce a variety of devices to detect urban change based on traces through the rank-size phase space, trajectories using a rank-time clock, and the definition of urban half-lives. We set this analysis within the wider context of stochastic simulation that is currently dominating discussion of scaling processes such as these.

Keywords: Emergence, Scaling, Rank-Size, Rank Clocks, Half-Lives, US Cities1790-2000

> "I will tell the story as I go along of small cities no less than of great. Most of those that were great once are small today; and those that in my own lifetime have grown to greatness, were small enough in the old days".

(from ***Herodotus***, the introductory quote in Jacobs, 1969)

1. INTRODUCTION

Emergence is a phenomenon which is intrinsic to the way systems grow and evolve. Biological systems, for example, start small from some seed, and as they change, novel and often surprising characteristics emerge which are clearly embodied in the basic rules enabling their cells to support growth. Where evolution is the focus, change usually implies growth although occasionally systems get smaller, often declining and this too sometimes leads to surprising features which are said to be emergent. Cities however have not been formally studied, until quite recently, as emergent phenomena although a moment's reflection suggests that there is no 'hidden hand' guiding their growth, and no top-down model for their physical and spatial organization. This is despite a series of highly suggestive works, some like Geddes (1915, 1949) **Cities in Evolution** first written nearly 100 years old, others of more contemporary origin such as Dendrinos's and Mullaly's **Urban Evolution** (1985), all of which attempt to articulate urban growth and form in terms of such change. Only recently as part of the growth of complexity theory which in turn is the modern expression of general systems theory, has our attention turned to how cities grow and evolve, to urban dynamics, and to how the patterns that we observe with respect to urban form and structure, emerge from a myriad of decisions from the bottom-up.

If our first theme in this paper is emergence, then our second is city size. Our concern here is not simply that cities emerge from the bottom-up which is clear enough. It is that systems of cities display a regularity that persists through time. This regularity which is an example of one of the many scaling laws that characterize different features in systems like cities, is itself a signature of emergent phenomena. This 'law' called variously the rank-size rule, Zipf's law, or Pareto's law, masks a complexity that persuades us into thinking that cities are rather simple sets of objects whose size and distribution simply reflects the hierarchy of size that is consistent with systems that work on fixed resources. However the fact that such overall relationships persist but the cities comprising them change their position in the hierarchy quite radically over short periods of time as Herodotus's quote above implies, is evidence enough that the way cities grow and evolve is far from simple. Indeed a significant issue is the emergence of new cities and the decline of existing ones but within the framework of this apparent aggregate stability. In this paper, we will outline the problem and present a preliminary approach that takes the matter a little further than that which has dominated our thinking about urban dynamics for the last 50 years.

The key issue in all this is the way we represent cities as distinct entities. The very idea of a distribution of sizes assumes that the objects comprising such distributions are themselves well-defined, unambiguous with respect to the way they can be measured. Cities might have been like this in historical times but their physical definition has collapsed in terms of their strict demarcation as the world has urbanized over the last 200 years. This suggests that there are at least two ways of representing size distributions: first in the traditional way which we will follow here by identifying what we still regard as individual cities, as much defined by our perception of locational distinctiveness and the way they are administered as by the fact that they have distinct boundaries; the second by taking the entire system as urbanized and partitioning it into more arbitrary units each with a degree of urbanization which enables us to examine the distribution of the size of these units and how they change. This is the way we previously examined the UK urban system (Batty and Shiode, 2003) but the problem here is that this is not appropriate to defining how 'new units' – new cities – emerge over time because all units are cities or parts of one greater whole which is the urban system. We will come back to this issue as we develop our argument but it is one of the central problems in a theory of cities where we live in an age where everything can be seen as 'city'.

We will develop this argument first by outlining the conventional wisdom of how cities emerge, emphasizing the mechanics of emergence in which positive feedback and historical path dependence are key features. We will then introduce the idea of city size distributions, sketching the rank-size relationship and the way researchers have dealt with its dynamics through changing relationships through time. This sets the scene for our analysis of the way the US urban system has evolved since 1790. We take data for the largest 100 town and cities at each 10 year time slice based on the decennial census, thus examining the stability of this relationship in terms of rank-size and the volatility of the distributions of individual cities within these distributions. We develop a number of ways of looking at these trajectories and conclude with a measure of the 'half-life' of cities which implies an index of the rate of change. We then note some ideas as to how we might explain such dynamics through simulation, linking our work to the massive body of research on how scaling distributions are formed and the stochastic processes that underpin to their structure. We conclude by arguing that although this area has been mined for the last half century or more, perhaps longer, there is much to do and the kinds of aggregate data that we use still contains much of interest in developing appropriate theories of urban dynamics.

2. THE MECHANICS OF EMERGENCE

The conventional wisdom about the long term evolution of world economy and society is that agriculture came first, then there were cities: in short that cities cannot get started until there is an agricultural base on which they can be sustained which in turn evolves from more nomadic pursuits (Childe, 1963). In fact, this view is now highly questionable. Jane Jacobs (1969) argues that urban pursuits do not depend upon the rural economy and that agriculture often exists within cities, as part of cities, and that innovations in agriculture certainly emerge from cities. The notion that agriculture leads to cities is too simple in that what is more likely is that there is continuing symbiosis between each which means that the rural is not prior to the urban, despite the fact that this is a deep seated assumption within contemporary society.

The other myth that is also deep seated is that cities are somehow economically advantageous in that they admit economies of scale. In fact there is considerable evidence that shows that rates of economic growth do not increase as the size of cities increases. In fact for the last 200 years or so in Britain, there is no relationship between the size of a city and its rate of growth, although there is greater variation in growth rates as city size gets smaller (Robson, 1973). This is quite consistent with the idea that the greatest volatility in growth occurs at the bottom of the urban size hierarchy in that this is where the process of weeding out places that are likely to become big and those that will not survive really takes place. There is a growing literature on rates of growth in cities which needs to be linked to the rank-size analysis. In essence, this implies that cities do not grow big because of any inherent economic advantage over small cities in terms of the rates at which they grow although once a city gets big, it acquires functions that imply some form of lock-in, or monopoly position. It is perhaps this that reduces the variation in growth rates for bigger cities.

In complex systems, objects change through a process of positive feedback in which their size is a function of their growth rate at any point in time. Assuming a constant rate of change, this is the process that leads to exponential growth (or decline) but there is nothing magical about this for it is simply a consequence of a growing population which is clustered. There may be limits on size due to capacity (or density) but local diffusion of population simply expands the object in its neighborhood and as a consequence the scale of the object changes. The key issue is that exponential growth does not depend on these growth rates of cities increasing as the cities themselves become bigger: this would lead to double exponential growth which is not an observed characteristic of cities in the past or present. This process of growth and decline however is complicated by historical inertia. Often the seeds that start this process of urbanization are planted in somewhat random fashion, despite there being good local reasons for the specific location of a place such as on a river or at trading post or on the edge of a lake. This leads to path dependence in that once a place gets started and then gets bigger, it is harder for it to become smaller in quite the same way as smaller places are able.

A simple but plausible model of the evolution of the US urban system for example which mirrors these conditions might be as follows: explorers and settlers from the 'old world' reached the eastern seaboard some 500 or so years ago at random points, determined by such events as local tides, winds, and so on. This in turn dictated where the initial places were established, mainly in Virginia, New York and New England. The places that grew were New York, Philadelphia, Boston, Baltimore, and the area south and west of present day Washington DC. Local diffusion and ad hoc decisions such as the founding of the capital led to a general growth westwards but the seeds that were planted first in the mid-west, then California, and finally the south west were determined too by local conditions and indeed by past historical settlement by the French and the Spanish. But what is so remarkable about this picture is that the initial seeds still remain. Our analysis below bears this out despite considerable volatility in terms of changing sizes and positions in the urban hierarchy since the late 18th century. This is a picture of strong path dependence – of aggregate patterns being somewhat inert but with much greater change occurring as one descends the urban hierarchy. To illustrate again such persistence, there is wonderful quote by Holmes (1992) in **The Oxford History of Medieval Europe** where he says: "Most Europeans live in towns and villages which existed in the lifetime of St. Thomas Aquinas, many of them in the shadow of churches built in the 13th century. That simple physical identity is the mark of a deeper continuity" (page iii).

It would seem that our argument is not so much that urban systems are marked by strong emergence in terms of city sizes but by inertia and persistence. In fact, the conundrum is to explain how particular small places become big within the more general pattern of spatial diffusion which occurs on all scales. This occurs as we shall see, within a strongly stable aggregate distribution of city sizes which might appear to act as a straight jacket on how the city system develops were it not for the fact that the aggregate pattern is itself an emergent phenomena: no one planned it. It is in this sense that we speak of emergence. We must now present this argument summarizing, albeit briefly, what is known about such size distributions and their historical evolution.

3. THE DYNAMICS OF CITY SIZE DISTRIBUTIONS

Zipf (1949) first popularized the fact that if one ranks cities in order of their size, the cities follow a regular inverse function – a rectangular hyperbola – a special case of the inverse power function $x^{-\alpha}$ where α is unity. Although he demonstrated this for a variety of city size distributions in different countries, this fact had been known previously by others: Auerbach and Lotka for cities, Pareto for income distributions, and even Max Weber had noted this and there is some suggestion that the Physiocrats had also been aware of this. Probably Leonardo had known of it and if this were the case, then so did the Greeks, taking us back full circle to Herodotus and the quote that introduces this paper. More significantly, such relationships are scaling, implying self-similarity which is the signature of processes that operate on all scales and give rise to fractals and all that this implies. They have become of central importance to theories of complex systems formed by processes operating from the bottom up, and currently this is where all the action is on the rank-size rule, as it came to be called.

In its strong form, Zipf's law says that the size of a city of rank r, called P_r, in the set of cities under consideration (often referred to as the hierarchy) is given by the size of the largest city P_1 divided by its rank. That is, $P_r = P_1/r$ which in more general form can be written as $P_r = K\,r^{-\alpha}$ where Zipf argued that $\alpha \approx 1$ and $K \approx P_1$. On double log paper, this relationship yields a straight-line and this is usually the way it has been estimated traditionally. In fact the fit to the top level of cities in many countries is so good that this has sustained interest in the notion that there is a 'law' to be discovered. Recently Krugman (1996) said: "We are unused to seeing regularities this exact … it is so exact that I find it spooky. The picture gets even spookier when you find out that the relationship is not something new – indeed the rank-size rule seems to have applied to US cities at least since 1890!" (page 40). As we will show in our reworking of Zipf in this paper, this stability and exactness is even longer lived.

The traditional explanation by Zipf (1949) is widely regarded as being somewhat mystical being based on intuitive notions of how opposing forces – unification and diversification – 'lead' to the relationship. In fact he does not show how this occurs although a number of geographers have illustrated how the rank-size rule, in its weak form as a power law, is consistent with central place theory. Rapoport (1968) in a thinly disguised critique of the law, poses the central issue when he says: "Clearly, if objects can be arranged according to size, beginning with the largest, some monotonically decreasing curve will describe the data. The fact that many of these curves are fairly well approximated by hyperbolas proves nothing .... No theoretical conclusion can be draw for the fact that many (reverse) J curves look alike. Theoretical conclusions can be drawn only if a rationale can be proposed that implies that such curves *must* belong to a certain class."

In fact, Zipf's (1949) work despite its influence is riddled with inconsistencies. He refers to the relationship as the rank-frequency, although rank is clearly cumulative frequency while frequency is size only in the sense that each individual making up the object – in this case each individual in the population of the city – is counted thus. More recent work however has sought to clarify this and there is now some consensus at least in statistical physics and economics concerning ways in which these distributions should be characterized (Adamic, 1999). This focus on the statistics of these distributions has given rise to the most promising theories to date although these are largely concerned with finding the underlying stochastic mechanism generating the probability distributions in question. Simon (1955) was the first to do this formally although Gibrat (1931) is credited with the underlying argument which he developed for income distributions, although noting this also applied to city sizes.

In essence, this stream of work which is now dominant, considers the rank-size relationship as being consistent with a process that generates not a power law for frequencies of cities by size but a log normal distribution. The long tail of the log normal, sometimes called the fat tail of its cumulative distribution which is the rank-size, can be approximated by an inverse power law and it is this that is considered as being the dominant data describing city sizes. This raises the problem of defining where a city begins in that at the bottom end of the lognormal, individual populations are the objects in question; these are judged to be qualitatively different from cities and thus there has been considerable effort in trying to devise generating mechanisms – with some success it might be noted – which cut-off the short tail using various thresholding arguments. We will not refer to this work in detail here for this paper is largely about empirical analysis where this work acts as background. But the reader should be referred to three key works that indicate various concerns from different disciplinary perspectives: from economics, see Gabaix (1999), from statistical physics, see Blank and Solomon (2001), and for a general review of all these fields, see Li (1999).

The way we will proceed here is to follow the emerging conventional wisdom. We assume that the most parsimonious models to date are those that are based on the random incremental but proportionate law of growth which is consistent with there being no increasing returns to scale in terms of the growth rate. These are based on Gibrat's Law of Proportionate Effect which generates the lognormal. Here we will not be concerned with what happens at the bottom of this distribution in terms of where new cities come from but we will simply take the top 100 cities (for the US urban system) and examine how these change., This means that we can quite safely measure the stability of these city size distributions using the rank-size rule and this is consistent with our concern for what happens as new cities enter and leave the top 100. However in a later paper we will update our analysis by examining models that simulate distributions from which we can measure the top 100 cities and this will involve us in a more comprehensive analysis of urban change at all levels of size. We have already attempted this for the UK urban system but there we assumed that the space economy was exhaustively subdivided into a fixed number of urban spaces – cities? – whose dynamics we then proceeded to analyze comprehensively (Batty, 2001). Only indirectly however do we address the problem of where new cities come from in this analysis; our concern here is to provide a new glimpse of the empirical dynamics operating in the US urban system as a first shot in devising better theory which we hope will follow.

In fact, Zipf (1949) did examine the temporal dynamics of the US system in that he plotted the rank-size distributions for all cities greater than 2500 persons from 1790 to 1930. He did not fit his law to these curves although it is fairly clear visually from his analysis that his law $P_r(t) \approx P_1(t)/r$, $t = 1790,...,1930$ is borne out. We reproduce his plots in Figure 1 below but it is worth noting that he only estimated two relationships formally – both for US metropolitan areas greater than 50000 population in 1940: for the top 100, he estimated $\log P_r = -0.983 \log r + 7.050$ and for the top 140, the same relation is given as $\log P_r = -1.036 \log r + 7.112$. We will rework and extend his analysis in the treatment that follows.

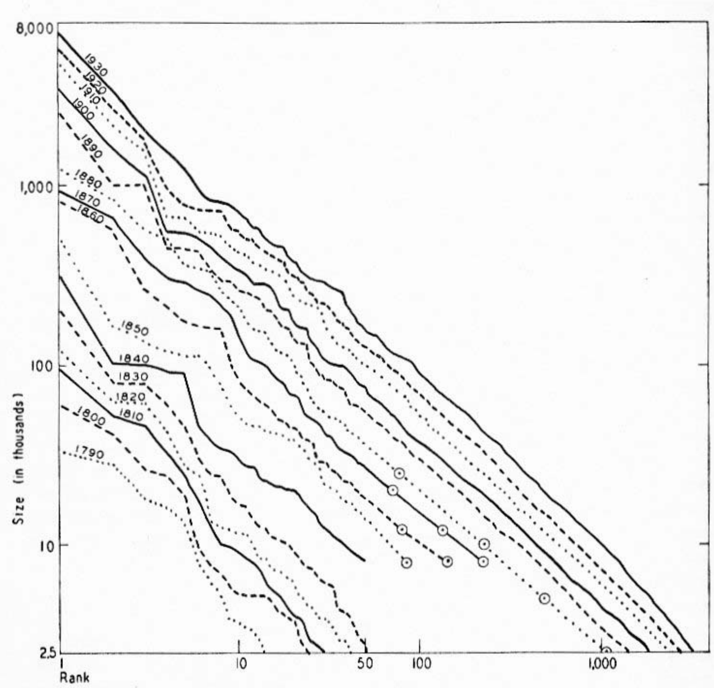

Fig. 10–2. U. S. A. 1790–1930. Communities of 2500 or more inhabitants ranked in the decreasing order of population size.

Figure 1 The Baseline: Zipf's (1949) Analysis of US City Size from 1790 – 1930

4. THE EVOLUTION OF THE US URBAN SYSTEM, 1790 TO 2000

4.1 The Basic City-Size Relationships

The data we have used is based on the top 100 population sizes defined for incorporated places in common usage in the US Census and as assembled by Gibson (1998) from 1790 to 1990. We have added to this the 2000 data which we consider to be generally similar, notwithstanding a couple of minor definitional differences. These places represent the most minimalist of city definitions in that they represent the first places that are defined as such by their populations changing slowly in definition during the 210 years during which we conduct our analysis. For example, New York City is essentially Manhattan until 1890 when the Bronx, Brooklyn, Queens and Staten Island are then added. We have to live with such changing definitions although we do consider this kind of change to be one of scale change in that the very definition of what constitutes a city changes radically during this period. It might be said that the objects in question change qualitatively during the period and our definition does not try to discount this.

There are many other definitions of US cities by the Census Bureau itself but these have not been standardized in an appropriate series. Although of great relevance to the consistency of these city size relationships across different scales, connecting our analysis to these must await further standardization which we intend as part of future research. Tom Wagner of the University of Michigan is working on this problem and some preliminary results can be seen at http://chinadatacenter.org/Presentation/Tom_zipf.htm . Our analysis here in fact is closest to that undertaken by Zipf for essentially the data between 1790 and 19390 that he used is taken from the same units that we use here.

In Figure 2, we show plot the basic relations on log-log paper which provides a direct comparator to Figure 1. The strength of the relationship over time is remarkable with the most obvious pattern being in the growth of the population in exorable fashion as the average size of the top 100 cities increases. In fact, the total number of cities over 2500 does not reach 100 until 1840, the numbers from 1790 to 1830 being 24, 33, 46, 61, and 90. The average size of city population in the set changes from 8402 in 1790 to 568698 in 2000. It is not possible to see any substantial changes in pattern from these graphs which plot cities in rank-size space. It is hard for example to see the opening up of the mid-west, the far west and the south west as a distinct pattern shift in these relations. What is clear is that the decision to impose an arbitrary cut-off at 100 cities does reveal that we are only examining the tip of the iceberg with respect to the entire urban system although this is essential for our emphasis on the entry and departure of cities from the top 100. During this period, the population of the top 100 cities has grown dramatically and we show this in Figure 3 where we have plotted the total population $P(t) = \sum_r P_r(t), \forall t$, and the population of New York City, $P_1(t), \forall t$, the number 1 ranked population center throughout the period which acts, as we shall see, as a kind of anchor point to the system. In fact the largest cities in the data set are capacitated in that like New York City, they show logistic growth which reflects density limits despite a recent surge (1990-2000) reflecting some oscillation around these limits.

Our estimates of the relations in Figure 2 reveal very strong correlations and remarkable consistency in the parameters α and K with these values being close to the strong Zipf values of 1 and P_1 at each time. We show these in Table 1. However there is a slow but sure drop in the value of the Zipf exponent α which is indicative of a system that is becoming less concentrated through time as reflected in the fact that the biggest cities remain the biggest but increasingly reach their capacities. This appears to have accelerated since 1960. To generalize these interpretations, we would need to rework this same analysis at different scales with different definitions of cities. In fact it is instructive to look at the intercept which is the predicted population of New York City. We have graphed the predicted and observed values of this population $P_1'(t)$ and $P_1(t)$ in Figure 4 where it is quite clear that although the model over predicts the population from the mid to the late 19th century, once the automobile society became established from the early 20th century on, there is systematic under prediction of this value. From Table 1, the correlations of observed with predicted for the whole set of cities are unerringly high but when we examine the deviation between predicted and observed values for the largest object in the set – New York City – which we define as $\Gamma(t) = 50\left[\left(P_1'(t) - P_1(t)\right)/P_1(t)\right]$, we see from Figure 4 that there are substantial deviations of up to 20 percent. This provides a somewhat different perspective on the quality of the fitted rank-size relationships.

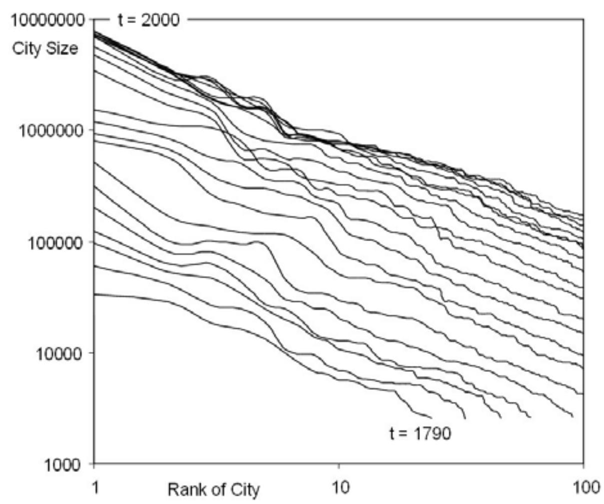

Figure 2 Rank-Size Space: Zipf Plots for US Cities from 1790 to 2000

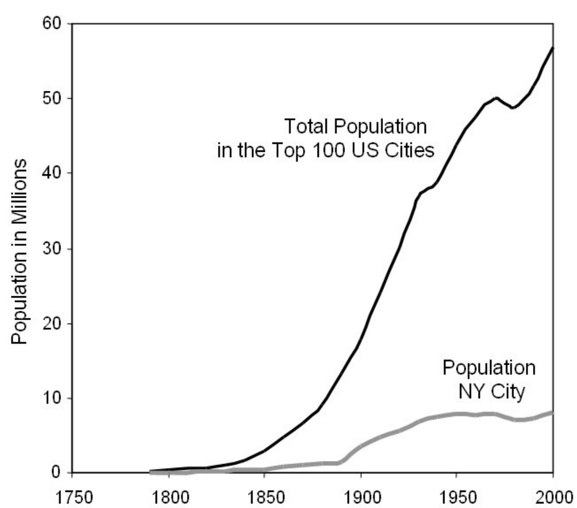

Figure 3 Growth of Total Population of the Top 100 Cities from 1790 to 2000

Table 1 Rank-Size Parameters for the Evolution of the US Urban System from 1790 to 2000

Year	r-squared	Intercept $\log K = \log P_1'$	Zipf exponent α
1790	0.975	4.660	0.876
1800	0.968	4.797	0.869
1810	0.989	5.005	0.909
1820	0.983	5.080	0.904
1830	0.990	5.244	0.899
1840	0.991	5.416	0.894
1850	0.989	5.689	0.917
1860	0.994	5.943	0.990
1870	0.992	6.102	0.978
1880	0.992	6.244	0.983
1890	0.992	6.384	0.951
1900	0.994	6.477	0.946
1910	0.991	6.568	0.912
1920	0.995	6.668	0.908
1930	0.995	6.750	0.903
1940	0.994	6.776	0.907
1950	0.990	6.827	0.900
1960	0.985	6.799	0.838
1970	0.980	6.789	0.808
1980	0.986	6.734	0.769
1990	0.987	6.730	0.744
2000	0.988	6.763	0.737

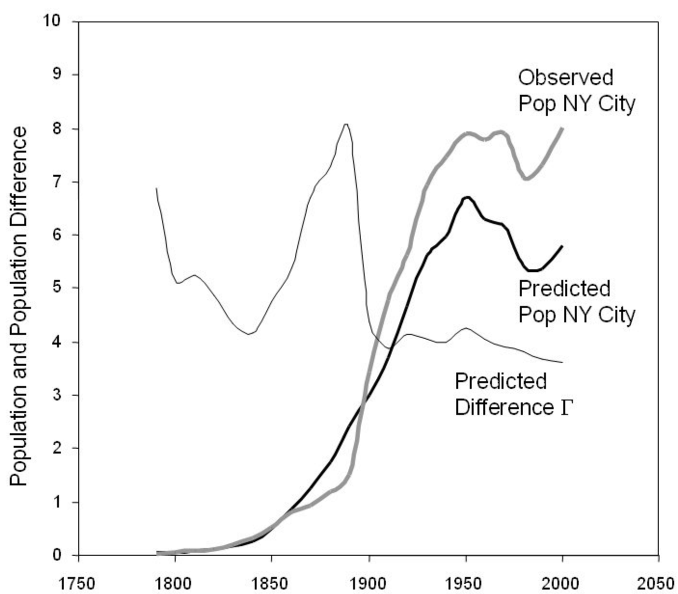

Figure 4 The Population of New York City, the Top Ranked City over 210 Years

4.2 Trajectories of Key Cities in Rank-Size and Rank-Time Space

There are some 267 distinct cities which appear in the top 100 over the 210 year period and the best way to examine these is to take a representative sample based on the dominant towns at the beginning of the period (1790), at the end (2000) and those towns which remain in all the sets throughout the period. The top 5 towns in 1790 are Baltimore, Boston, Charleston, New York, and Philadelphia while in 2000, the top 5 are Chicago, Houston, Los Angeles, New York and Philadelphia. This clearly shows how the mid west, west and south west have opened up during the period. The other two towns that have remained in the list throughout the period really represent the south and have been supplanted now by Washington DC. These are Norfolk and Richmond, both in Virginia. Charleston however is the only town in this overall list of 10 which drops off the list, disappearing from the top 100 in 1910.

Our subsequent analysis suggests that we can classify these 10 towns into five distinct groups. New York City is outstandingly different being the anchor for the entire system and unwavering in its position and dominance as is evidenced by many other considerations which are not part of the analysis. The second class is those eastern cities which were dominant in the late 18th century and have remained dominant but whose position has been eroded by the opening up of the rest of the country and by suburbanization and sprawl: these are Baltimore, Boston, and Philadelphia. The third group is the cities which enter the list later as population growth in the US has spread west and these are Chicago, Houston and Los Angeles. The fourth group and perhaps also the fifth are the colonial cities of south: the fourth are the two cities of Richmond and Norfolk in Virginia which generally decline in rank. Indeed Richmond looks as though it will fall out of the list by 2010 although Norfolk is probably the recipient of greater growth due to the location of the Navy and the generally favorable climate. Charleston, our fifth city, is a classic colonial town which disappears as soon as the US economy begins to spread west and industrialize.

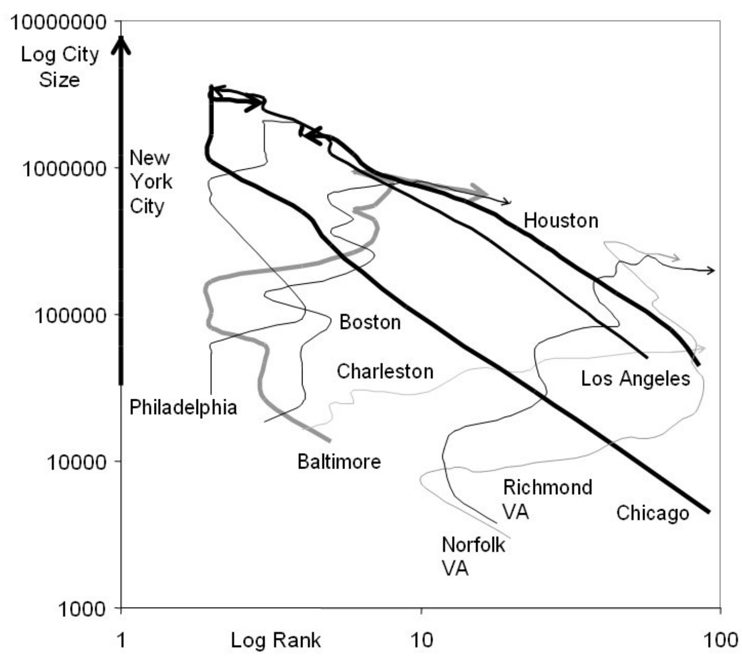

Figure 5 Traces of Ten Key Cities in Rank-Size Space

These features are very clear from the two phase space diagrams in which changes in their rank and size through time can be illustrated. First in Figure 5, we show the trajectory of each city in rank-size space. This is equivalent to tracing each city in Figure 2 and plotting its position solely with reference to itself and not with reference to the entire rank-size distribution. New York City is clearly an outlier but the new cities of Chicago, Houston and Los Angeles grow in rank status and in population through the period, moving from the bottom right hand side of the space to the top left. The colonial towns tend to move towards the edge of the phase space while the eastern seaboard cities retain their approximate rank up until the last 50 years but then begin to lose status while continuing to grow in population. These cities head off from the bottom to the top of the space. The biggest cities also lose population in the late 20^{th} century although this is hard to see clearly in the diagram.

Another way of examining these changes is to throw out size and simply concentrate on rank with time being fashioned using a different geometry. What we have done is to construct a rank-time clock which arranges the 22 time slices from 1790 to 2000 in a circular form reminiscent of a clock where movement through the decades is in the clockwise direction. Here we plot the rank of the city through time. A city which remains at a given rank traces a perfectly circular orbit while one which disappears from the space moves off in a clockwise spiral from a more central to a less central location. The colonial cities fall into this category while New York City's rank at 1 places it always at the centre of the circle. A city which moves up the hierarchy over time also traces out a spiral trajectory but with the movement from the edge of the space toward the centre. We illustrate these trajectories for all 10 cities in Figure 6. In 6(a), we show the pattern using a clock where rank is in absolute terms. As those ranks which are highest are compressed to orbits at the center of the clock, and as we are selecting cities which are dominant, to see the action towards the centre, we need to stretch these orbits and accordingly we illustrate the same data in Figure 6(b) where we have used a logarithmic scale for the rank. Here is it quite clear for example that Philadelphia which can barely be seen in Figure 6(a) traces a more or less circular path over the entire period in that it starts and ends near the top of the list.

This clock enables us to examine the time at which new cities enter the space and the time at which old cities leave. It also enables us to work out the speed at which a city is rising up the hierarchy or descending the same. Cities which are unchanging trace circular orbits or as in the case of New York City collapse into the centre as a point. Cities that more gently come into the space or leave it are more likely to trace regular spirals while cities that come in rapidly and stay in position or leave in the same way trace out lines which are straighter than either the spirals or the circular orbits. In this way we might be able to develop a geometric classification of cities without needing to plot such diagrams and in this way examine the entire set of cities in the top 100 lists for each time slice. This is something that would be worthwhile in that we would then begin to see what the mix of city types was in terms of their dynamics within the entire set rather than as we have done here, restrict our analysis to a sample that marks those which dominate the US urban system at the beginning and the end of the period as well as those that are most persistent.

4.3 The Half Life of US Cities

Our last foray into this data set begins to examine for the first time, the extent to which new cities enter the list at each time slice and the extent to which cities already in the list leave the list. To do this for any time slice t, we compute the number of cities which are in the current list at time t, are in a previous list at time $t-n$, and in a future list at time $t+n$. We call

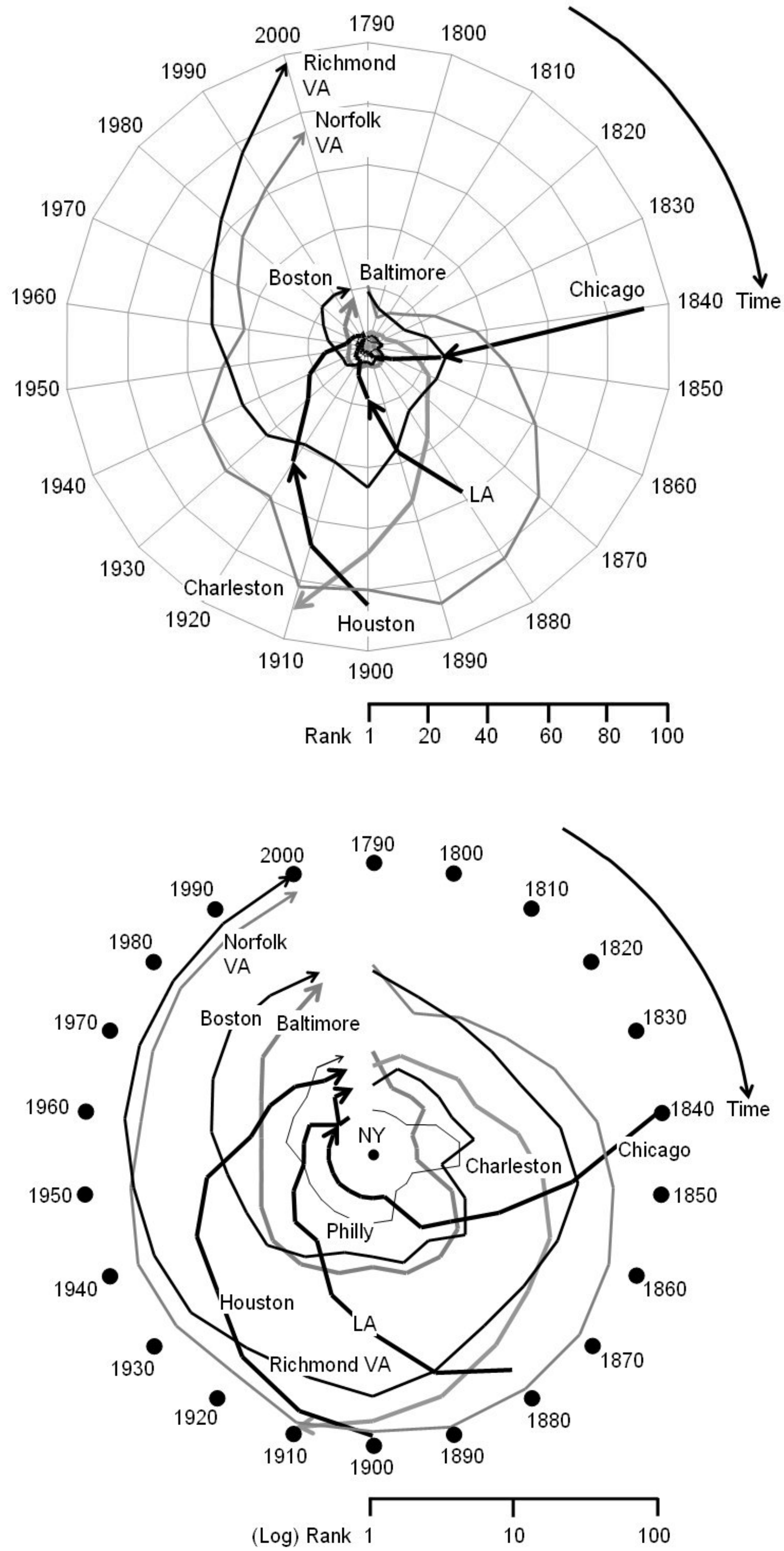

Figure 6 Trajectories of the Ten Key Cities on the Rank-Time Clock in Absolute and Logarithmic Space

this number $N_{t,t\pm n}$. In general assuming that we are not at the beginning of the time period when there are less than 100 cities in the list, that is assuming we are dealing with times from $t = 1840$, then it is clear that those cities in the list at time t, are always $N_{t,t} = 100$; as the time from this date increases or decreases, then in general we will have $N_{t,t\pm n} \leq 100, n > 0$. We can thus form a matrix where time runs from $t = 1790$ to $t = 2000$ which shows the number of cities at any time $t \pm n$ associated with those comprising the list at any time t. From this we are able to compute a half life of cities. In this paper, we will not pursue this rigorously but simply sketch the idea from the appropriate graphics.

We plot these numbers of cities for each time t in Figure 7. What we see here is a typical pattern of decay in terms of the number of common cities with respect to the list as time changes away from the date in question in either direction – before or after. To give an example, if we take the 100 top ranked cities in 1900, then in 1910 only 90 of these cities are still in the list, a loss of 10, in 1920 only 84, a further loss of 6. By 2000, only 47 of these cities remain and if the half life is the time it takes to retain one half the original list, then this would actually be 1990 when exactly 50 cities remain. In the other direction, in 1890, 93 cities are in the list, in 1880 82 cities and back in 1790 only 11 cities. The half life in this direction is somewhere between 1850 and 1860. In this sense the backwards half life – around 50 years - is much less than the forwards half life – which is around 90 years but this is because we are dealing with a rapidly growing system which we break into in its growth phase.

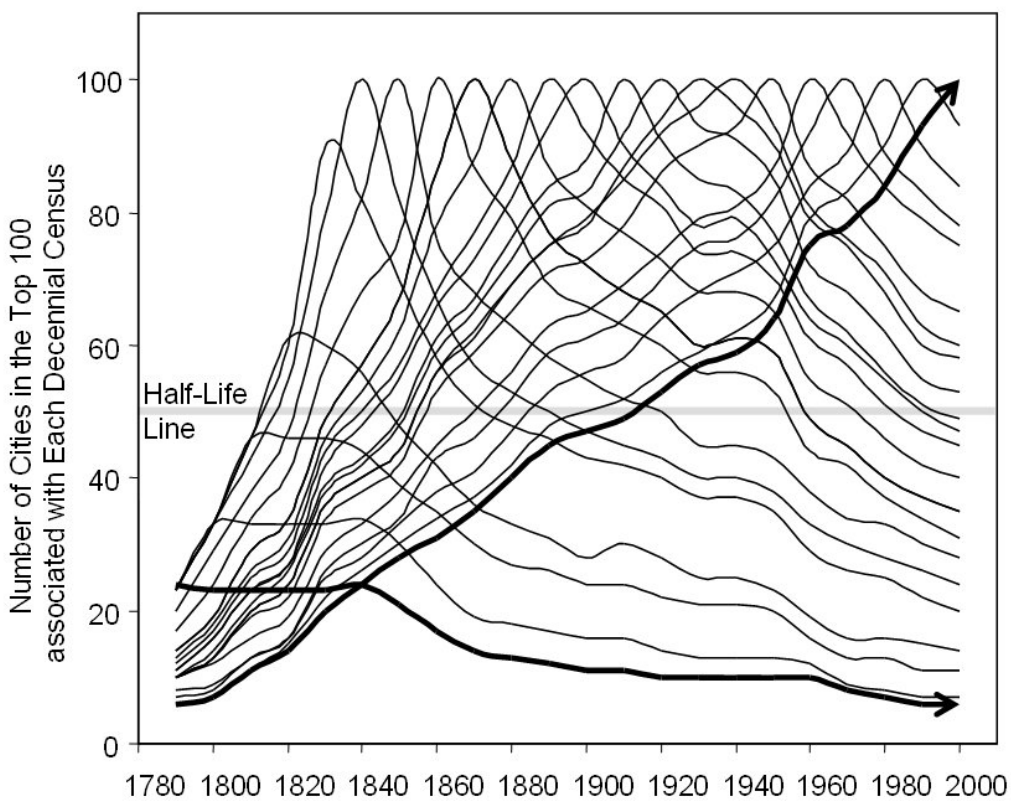

Figure 7 The Lives of US Cities from 1790 to 2000

We can actually compute these half lives quite accurately using equations such as that for the forward half life $\sum_n N_{t,t+n} n \Big/ \sum_n N_{t,t+n}$. In this paper, all we will do is to indicate that we can get an estimate of the half life for each time period if we simply examine the intersection of the curve formed from $N_{t,t\pm n}$ with $N = 50$, thus reading off the dates $t-n$ and $t+n'$ which give estimates of the backwards half life n and the forwards half life n'. We show this line on Figure 7. We can also detect from Figure 7 the fact that the system is quite asymmetric in terms of its growth profile. Cities enter at a faster rate in general than those leaving as the system builds up in the early 19th century. In the 20th century, the depression and war years 1920-1950 seem to suggest that there is less growth and thus less change with respect to cities in the top 100 while in the last 40 years there appears to have been an acceleration of cities both entering and leaving the top lists. There is much more we might say about these dynamics but this will require more detailed analysis. For now we rest content to simply indicate the way this analysis might be taken forward.

5. CONCLUSIONS: SIMULATING RANK-SIZE DISTRIBUTIONS

Although we have shown that despite the apparent stability of the urban system as evidenced through the rank-size relationship, there is considerable volatility in that the composition of cities comprising these relationships which change almost entirely over a 200 year period. Indeed although the forward half life of cities is about 100 years, the backwards half life reflecting the speed at which new cities are created (in this sense enter the top 100 ranks) is about 50 years. We have shown that we can develop quite detailed analysis at this macro level which has interesting and important implications for a theory of urban dynamics but to progress any further we need to consider how we might engage in simulation.

We have already attempted some rudimentary simulation based on the stochastic proportionate effects and examined the degree of volatility of such systems with respect to the entry and exit of new cities (Batty, 2001). However most of the models which attempt such simulation are somewhat artificial in that they do not incorporate any competition between the cities: cities exist independently of one another as for example in the simulations developed by Blank and Solomon (2001). There are some useful extensions to these models, for example that developed by Manrubia and Zanette (1998) but even in this case where local diffusion is used to spread development, there is no real competition built into the framework. A more important limitation however is that there is no inertia in these stochastic models. For example although the US urban system displays the same degree of volatility in terms of new cities entering and old leaving as those stochastic models, it is not possible to replicate the existence of cities that remain in place. New York, for example, has remained the dominant city throughout American history and despite pronouncements of its imminent demise, particularly since 9/11, this analysis suggests that its role in the system is fundamental and deep seated and that it is unlikely to lose this position in the foreseeable future. This kind of inertia is hard to replicate in stochastic models as it may well relate not simply to the internal dynamics of the US system but to its external dynamics, to its role as a world city and as an anchor point back to the old world from which America was originally spawned.

It is not easy either to simulate the opening up of an urban space, in other words to simulate colonization of empty space in the way it occurred in America with a prior but very loosely developed settlement structure already in place. This really gets to the heart of the matter because there are still no models to date which actually create new events in the way new cities are clearly invented. What we need is some form of growth model in which tension in the system creates the opportunity for new events and new events – cities – emerge to fill this niche. The way we have attempted this so far is to assume that such events already exist – that locations exist but they are not active in that they do not contain activities. In a sense this is unsatisfactory because the spaces already exist as dormant events or objects. Moreover this is complicated too by the fact that the potential array of stochastic processes that in principle can generate these events, begin with objects which are of minimal size, single persons for example which are well below the threshold of what we would consider a city. All this analysis is thrown back once again onto the problem of what constitutes a city. A person is clearly not a city but for purposes of simulation it may be necessary to consider a much wider domain of settlements than has hitherto been the case in thinking about city size distributions. This paper has raised more problems than it has resolved but future research is clear: to extend this analysis to a wider range of city sizes and to see those that we consider 'proper' cities as some subset of these, both with respect to empirical observations of the real city system and potential models that might be used in their simulation.

ACKNOWLEDGEMENTS

Rui Carvalho (UCL), John Nystuen (U Michigan), Naru Shiode (SUNY-Buffalo), Tom Wagner (U Michigan) and Yichun Xie (Eastern Michigan) all contributed through discussion to these ideas. The ESRC Nexsus Project (L326-25-3048) provided partial support for this research.

This paper was first presented at **the 8th International Conference on Computers in Urban Planning and Urban Management** (CUPUM '03) at Sendai, Japan, May 27-May 29, 2003.

REFERENCES

Adamic, L. A. (1999) Zipf, Power-laws, and Pareto – A Ranking Tutorial, http://ginger.hpl.hp.com/shl/papers/ranking/ranking.html. Accessed 9/2/03

Batty, M. (2001) Polynucleated Urban Landscapes, **Urban Studies**, **38**, 635-655.

Batty, M. and Shiode, N. (2003) Population growth dynamics in cities, countries, and communication systems. In P. A. Longley and M. Batty (eds.), **Advanced Spatial Analysis**. ESRI Press, Redlands, CA, in the press.

Blank, A, Solomon, S. (2001) Power Laws and Cities Population, http://xxx.tau.ac.il/html/cond-mat/0003240. Accessed 9/2/2003.

Childe, G. (1963) **Social Evolution**, Fontana-Collins, London.

Dendrinos, D. with Mullaly, H. (1985) **Urban Evolution: Studies in the Mathematical Ecology of Cities**, Oxford University Press, New York.

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

Geddes, P. (1915, 1949) **Cities in Evolution**, Williams and Norgate, London.

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

Gibson, C. (1998) **Population of the 100 Largest Cities and Other Urban Places in the United States: 1790 to 1990**, Population Division Paper 27, US Bureau of the Census, Washington, DC; at http://www.census.gov/population/www/documentation/twps0027.html Accessed 14/4/03

Holmes, G. (1992) Preface. In G. Holmes (ed.), **The Oxford History of Medieval Europe**, Oxford University Press, Oxford, UK.

Jacobs, J. (1969) **The Economy of Cities**. Random House, New York.

Krugman, P. (1996) **The Self-Organizing Economy**, Blackwell, Cambridge, MA.

Li, W. (1999) Zipf's Law, http://linkage.rockefeller.edu/wli/zipf/. Accessed 27/4/2003

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

Rapoport, A. (1968) Rank-Size Relations, **International Encyclopedia of the Social Sciences**, **Volume 13**, 319-323. (Editor David L. Shills, The Macmillan Company and the Free Press, New York).

Robson, B. (1973) **Urban Growth: An Approach**, Methuen, London.

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

Zipf, G. K. (1949, 1965) **Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology**, Addison-Wesley and Hafner Publishing Company, New York.