CENTRE FOR ADVANCED SPATIAL ANALYSIS
Working Paper Series

CASA

Paper 19

POWER LAW DISTRIBUTIONS IN REAL AND VIRTUAL WORLDS

Narushige Shiode
Michael Batty

UCL

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

Tel: +44 (0) 20 7679 1782
Fax: +44 (0) 20 7813 2843
Email: casa@ucl.ac.uk
http://www.casa.ucl.ac.uk

http://www.casa.ucl.ac.uk/powerlaw.pdf

Date: February 2000

ISSN: 1467-1298

We are grateful to Martin Dodge (1999) who originally collected the data on Web size and hyperlinks from ***AltaVista***

ABSTRACT

This study compares the statistical patterns of size and connectivity of the global domains (as in ".com" and ".uk") to the geographical distribution of the global population. As the development of Web sites represents the cutting edge of the new global economy, their sizes and contents are likely to reflect the distribution of population and the urban geography of the real world. There is widespread evidence that population and other socio-economic activities at different scales are distributed according to the rank-size rule and that such scaling distributions are associated with systems that have matured or grown to a steady state where their growth rates do not depend upon scale. In this paper, we advance the hypothesis that the growth of Web pages in different domains is not yet stable. This is supported by our analysis that shows that the most mature domains with the most pages follow near rank-size relations but that countries which are much less advanced in their development and use of internet technologies show size relations which, although scaling, do not conform to rank-size. Our speculation is that as the Web develops, all domains will ultimately follow the same power laws as these technologies mature and adoption becomes more uniform. As yet we are unable to support our hypothesis with temporal data but the structure in the cross-sectional data we have collected, is consistent with a system which is rapidly changing and has not yet reached its steady state.

1. Introduction

Rapid deployment of information technologies and the exponential growth of the World Wide Web are beginning to generate a new geography within the wider structure of cyberspace (Batty 1993, Ludwig 1996). Various attempts at measuring and interpreting the structure, size and connectivity of this space have been made but its growth and evolution generate a constant need for new measurements and interpretations (Abraham 1996, Bray 1996, Pirolli *et al*. 1998, Pitkow 1998, Adamic 1999).

In general, as Web sites clearly form an integral part of social and economic development, their sizes and contents are likely to reflect the distribution of population and the urban geography of the real world (Gorman 1998, Mitchell 1999). Recently, it has been predicted that, despite its apparent arbitrariness, the sizes of Web sites and hyperlinks between them follow known distributions of growth phenomena such as those observed for cities and regions (Albert *et al*. 1999, Faloutsos *et al*. 1999, Huberman and Adamic 1999).

We begin by reviewing these recent investigations, and we then extend this to Web sites which are distributed geographically in real space. Through a comparative study of the sizes of global domains and national populations, we argue that the sizes and frequencies of Web sites follow those well-known scaling distributions first catalogued for a variety of different social phenomena by Zipf

(1949), and subsequently widely applied to city size, income, word frequency, and firm size distributions.

2. Related Studies of the World Wide Web

Although the Internet has only become significant during the past ten years, it has already attracted a number of researchers who have conducted various investigations and surveys of its distribution and size. A vast amount of statistical resources as well as numerous theoretical contributions to interpreting the growth of the internet in general and the Web in particular exist, and among the many studies conducted thus far, four approaches to Web analysis can be identified. We will list these by way of setting the context to our work.

> *Statistical approaches based on data summaries using inferential statistics*: the most obvious yet vital method for grasping the overall impression of the Internet is to collect its statistical information. A number of institutes have attempted to capture the state of the Web through a survey on the number of various Web sites, active servers, users of the Internet and the growth rate of each of these (Gray 1995, Bray 1996, Coffman and Odlyzko 1998, MIDS 1999, Network Wizards 1999, OCLC 1999). However, due to the exponential growth rate of the Internet and its increasingly complex structure, most of these figures inevitably consist of estimated values, or the rough indicators of its scale (Network Wizards 1999).
>
> *Visualisation based on the graphical representation of data:* most of the services provided by the Internet such as the World Wide Web are of metaphorical content and have no physical entity. Various cartographic and geo-information techniques are being applied to visualise this virtual domain from a variety of perspectives. Some focus on the pattern displayed by search queries (Carriere and Kazman 1999), while others depict the topological connectivity of hyperlinks (Shiode and Dodge 1999). Visualisation, if properly applied, can provide persuasive, intuitively comprehensible outputs. However, such approaches are usually self-conclusive and limit the possibility of further exploration of content.
>
> *Data mining based on low-level tools that uncover anomalies and clusters*: in contrast to the statistical approach, data mining typically focuses on a single local spot or on a particular point of interest and carries out in-depth analysis to comprehend the exact impacts and effects at lower levels. Examples include local traffic distance measurement (Murnion and Healey 1999) and IP address distribution at the district level of a country

(Shiode and Dodge 1998). The only limitation is that while such methods can be applied to a local or specific aspect of the Web, it is practically impossible to maintain the level of detail if the entire Web needs to be searched as we invariably wish.

Model-based approaches employing theoretical tools to interpret and simulate structure and growth: this final approach aims to understand the Internet by constructing a model of its structure. In particular, there is an extensive collection of studies on its connectivity and topological structure (Abraham 1996, Kleinberg 1997, Wheeler and O'Kelly 1999, Barabasi *et al*. 2000). Among these studies is the application of a social network concept which reflects the 'small world' assumption (Watts and Strogatz 1998). The underlying idea is that for a variety of global network phenomena, all objects or people are connected to one other within a chain of six acquaintances, which is popularly known as the 'six degrees of separation'. Albert *et al*. (1999) have applied this concept to measure the degree of connectivity of the Web, predicting that Web pages are separated by an average of '19 clicks'. This connectivity measurement is closely linked to the idea of power laws describing networks where "… the probability of finding documents with a large number of links is significant, as the network connectivity is dominated by highly connected Web pages". (Albert *et al*. 1999).

Based on this last approach, we will conduct a rank-size analysis of the global domain based on countries and Web page hyperlinks within and between them. We will then compare these distributions with conventional social and economic indices of the real world; namely, national population and real GDP. First however, we will explain the basis of the power laws we will use, noting their relationship to rapidly growing systems such as the Web that we seek to model.

3. Power Laws, Scaling and Rank-Size: Zipf's Law Revisited

Distributions in nature and economy which are composed of a large number of common events and a small number of rarer events often manifest a form of regularity in which the relationship of any event to any other in the distribution scales in a simple way. In essence, such distributions appear to arise through growth processes which do not favour the common or rare events and which involve random additions to the set of events or objects. Typically, the size of an event $P(x)$ scales with some property of the event x in the form $P(x) = Kx^a$ where K is a constant and a some parameter of the distribution. Such distributions are scaling in that the size of the event is proportional to the size of the property; that is, if the property grows by l, then the size scales as $P(l\ x) = K(l\ x)^a = l^a Kx^a = l^a P(x)$. From this it is clear that $P(l\ x)/P(x) = l^a$ which has a particularly simple form when $a = 1$. These relationships can be formulated either in simple frequency form or in cumulative frequency form, usually as a rank-size relationship, which is preferred in this case, when the focus is on the rarer or larger events that dominate the distribution.

The best known of these scaling laws is the rank-size rule which was first popularised by Zipf (1949) for cities, word frequencies, and income distributions. Zipf's Law, as it is called, has the general form $P(r) = Kr^{-q}$ where $P(r)$ is the size of the event, in the case of cities - the population, r is its rank in descending order of size where $P(r) > P(r+1)$; q is some parameter of the distribution and K is a scaling constant. Sometimes the relation is presented as $P(r)r^{q} = K$ for any r which implies some form of steady state consistent with the growth process. The relevance of such simple scaling to city-size distributions has been known for over 100 years. Auerbach (quoted in Carroll 1982) proposed that the exponent q was 1 in 1913, while Lotka (again in Carroll 1982) suggested that $q = 0.93$ in 1925. Zipf (1949) and many others since then (see Krugman 1996) have confirmed this 'iron law' of city sizes. The usual way of fitting such distributions to data (which we follow here) is to perform a linear regression of $\log[P(r)]$ on $\log[r]$ where the parameters $\log K$ and q are the slope and intercept of the curve $\log P(r) = \log K - q \log r$, respectively.

There is considerable debate as to whether the systems and their size distributions modelled with power laws of this form are best represented by such log-linear relations. In fact, the Yule and log-normal distributions generated by various growth models and even stretched exponential, parabolic fractal and related forms might be preferable for distributions with fat, heavy or long tails (Laherrera and Sornette 1998). Here, however, we will develop the rank-size model largely because it represents a first attack on the problem of q measuring the size of the Web, and there are good stochastic models that are consistent with the kinds of distributions that we observe. In particular, Simon (1957) has developed a growth model which is based on three assumptions that appear to fit many natural and social systems. First, new events or objects are created at a regular but random rate and of the smallest size. Second, the growth rate of all existing events is essentially random, and third, the rate is independent of the size of objects, but with average actual growth proportional to size. As the number of events grows, their distribution converges to the steady state $P(r) = Kr^{-q}$ with $q = 1/(1-p)$ where $p < 1$ is the average growth rate of events which in the steady state converges to zero. This is a very useful interpretation; when the growth rate is near to 1, it means high value of and hence, indicates that the system is in its immature early stages, akin to that, for example, associated with the Web. As we will show below, our null hypothesis is that the system is already in the steady state with $q \approx 1$ but that deviations from this (which we will see in the rank-size plots), will indicate how far different domains (countries) in the system are from the steady state.

We are also aware of several other models which might be as appropriate as the rank-size. Simon's (1957) model is indeed equivalent to those which generate the Yule and log-normal distributions where the short tail of the distribution does not accord to the rank-size relation. In fact, most applications of scaling laws to these kinds of distribution 'conveniently forget' the short tail, fitting the model to the long tail, on the assumption that the size of events has to pass a certain threshold before the maturity of rank-size takes effect. Our interpretation suggests that the Simon model is compatible with explaining the short tail as well, although we will only briefly explore this point in this introductory paper.

As far as we are aware, the strict rank-size rule has not been applied to the distribution of Web pages and their hyperlinks for different country domains. However, Albert *et al*. (1999) use pure scaling to measure the frequency distributions of the numbers of in-degrees and out-degrees of links from Web sites, with implied values of $q \approx 1.45$ and $q = 1.1$ respectively for the associated rank-size relations. Faloutsos *et al*. (1999) have examined out-degrees from internet domains at three points in time in 1997-1998, and show that the equivalent q exponent varies from 0.81 to 0.82 to 0.74 for the rank-size rule and from 1.15 to 1.16 to 1.20 for the same data fitted in its simple frequency form. However, because these contributions stress connectivity, both works are almost entirely associated with hyperlinks found between a subset of the Web which is, at one level, comparable to the air route network in the real world, as opposed to the Web sites being the equivalent of city sizes.

In fact we would argue that the fundamental concept of the power law performs at its best when ranking a non-directional, agglomerative or accumulative set of events (or objects) that are spatially dispersed over a certain area. This accords with the developments of scaling laws in physics as well as in biology. Moreover, in order to comprehend the Web in a geographical context, it is essential to compare various distribution patterns associated with the size of the Web with those of the real world. In this light, we will measure the size of domains at the global level as well as hyperlinks observed within and between them. We will then compare them with the distribution patterns of national population and GDP. This not only contextualises Web size with a real geography, but also helps further to ground the earlier results obtained by the Albert and Faloutsos groups.

1. Data Sources and Information

For this analysis, we obtained data for population, GDP, Web site size and hyperlinks, the full listing of which is given in *Appendix A*. Using the ***AltaVista*** search engine, we obtained the total number of Web pages registered under 180 global domains that represent a nation, region or a large set of organisations of similar characters (e.g. "mil" as in the US Military). At the same time, we obtained the number of hyperlinks within and between these domains. Real GDP in billions of $US for 1998 (at 1990 values) and total population for 1994 were taken from IMF World Outlook (IMF

1999) and the GIS package ***Map/Info Professional*** respectively. Although we initially obtained data set for 180 global domains, we immediately excluded some of the data, conducting all our analysis with 151 data points for the following reasons:

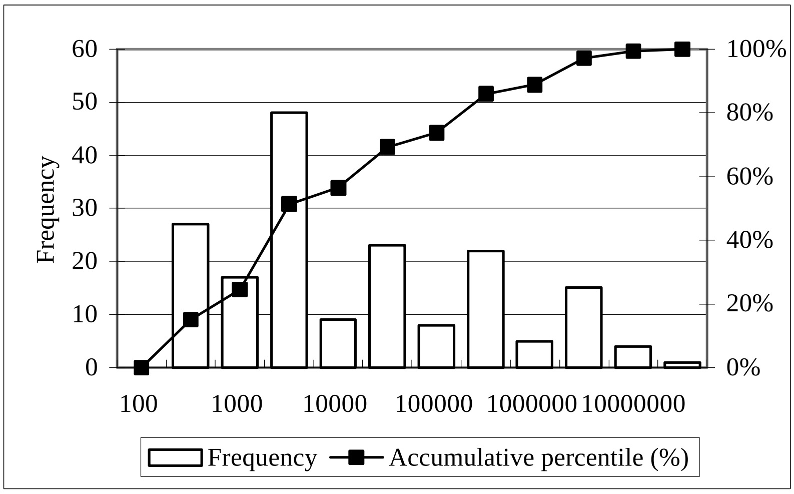
1. Consistent data could not be obtained for some countries and regions that have recently undergone radical transitions such as major political change or war (e.g. Hong Kong, Macedonia). The same applies to some regions with autonomous governments which are nonetheless part of other countries.

2. The breakdown for the non-regional domains such as “com”, “int”, “net” and “org” was difficult to estimate, and the ‘correct’ proportion could not be assigned to each participating country. There is some discussion that the US industries have up to 60% share of “com,” (Gray 1995) but the actual ratio remains uncertain, and breakdowns for the other three domains are unpredictable.

3. Considering the impact it has upon the entire rank, we included the US in our preliminary analysis, despite the above remark. We defined the US domain as a combination of the following: *United States* (us), American Samoa (as), Guam (gu), Puerto Rico (pr) and US Virgin Islands (vi); also *education* (edu), *government* (gov), *military* (mil) and 50% of *commercial* (com). As US firms may actually make up a significantly different percentage from 50% of “com” and other super-national domains, this introduces considerable uncertainty into the analysis. This ambiguity, however, would not have significant effects on rank-size analysis, if we were to focus on the rank size distribution below the second and lower rank domains. Abnormal values of highest ranked events are a common phenomenon. Primacy, in city size distribution, for example, has to be dealt with as reflected in large political and historical centres such as Paris, London and Berlin (Berry and Horton 1970), but in general, this does not lead to inconsistencies in rank-size *per se*.

4. We encountered a tremendous number of hits for countries with a peculiar domain suffix such as Columbia (.co) or Tonga (.to). We assume that this is caused by the purchase of these popular sounding domain names by the firms and individuals residing outside the country in question and this will certainly distort comparisons with the distribution of population as well as GDP. As yet we do not have a method for measuring the impacts of external contributors, and thus have to accept any such data at its face value. On the other hand, such peculiar agglomerations, especially those that differ from the GDP distribution pattern, may reflect a new geography of information space.

The domain size ranged from the super-scales of “com (*commercial*)”: 48,284,554 pages, and “net (*network*)”: 7,467,435; down to small country domains such as “cg (Congo)”: 109, and “tp (East Timor)”: 106. **Figure 1** presents a histogram of the domain sizes where over 25% of them fall within the intervals from 5,000 to 10,000 pages.



The Number of Domains in Each Scale Range		
0	Frequency	Cumulative
100	0	0.00 %
500	27	15.08 %
1000	17	24.58 %
5000	48	51.40 %
10000	9	56.42 %
50000	23	69.27 %
100000	8	73.74 %
500000	22	86.03 %
1000000	5	88.83 %
5000000	15	97.21 %
10000000	4	99.44 %
50000000	1	100.00 %

Figure 1. The Distribution of Domain Size

The number of links between the 180 global domains was also investigated. We used script commands for generating multiple queries, n^2 separate queries for n number of sites, and counted the number of hyperlinks between each sub-domain by applying the syntax “**+url: <sub-domain1>.uk +link: <sub-domain2>.uk**.” Within the 16,111 possible combinations, we observed a total of 76,735,152 links of which 16.1% (12,318,346 links) were found between “com” and “net.” Whether the database of ***AltaVista*** search engine actually reflects an unbiased sample of the Web sites or not remains an open question. Nevertheless, it is considered to be one of the most comprehensive index of Web pages publicly available (Sullivan 1999), containing over 150 million Web pages (as of 1 February 1999). Thus, we assume that the ***AltaVista*** data reflect the actual state of Web and can be relied upon.

Correlations between Web size and the total number of links assigned to domains regardless of direction (that is, both incoming and outgoing links), are shown in *Appendix B*, together with those based on population and GDP. It is not surprising to find an r^2 for Web size and hyperlinks of 97%, but this simply confirms consistency in the average number of links per page. The overall average was 3.92, much lower than the 7 obtained by Albert *et al*. (1999). This may be partly explained by the differences in the methods of data collection. Albert's group counted the number of pages at some specific sites such as those of their own research institutes as well as the White House whereas our data, while globally obtained, depends on a commercial search engine.

5. Analysis and Interpretations

We have ranked in descending order the Web site, demographic, and economic data. This is measured respectively by the number of Web sites for each domain, number of incoming links into each domain (in-degrees), number of outgoing links (out-degrees), total links associated with each domain (in-degrees and out-degrees and inter-domain links), real GDP in billions of dollars US, and national population. In **Figure 2**, we present a complete graphical analysis of this data, plotting the distributions on logarithmic scales, visually associating various data, and computing idealised and actual rank-size relations.

None of the distributions follow the classic linear rank-size form for all distributions are concave to the origin. The largest sizes do appear to conform to simple power laws but the smaller sizes would be radically underestimated using these power laws. It is immediately clear from this analysis that the distributions of population and GDP are much closer over their larger size range to rank-size than any of the Web data. The rank-size is classic for the population of the largest 100 or so countries (out of 151) with GDP the same for over half (75). We consider that the smaller than

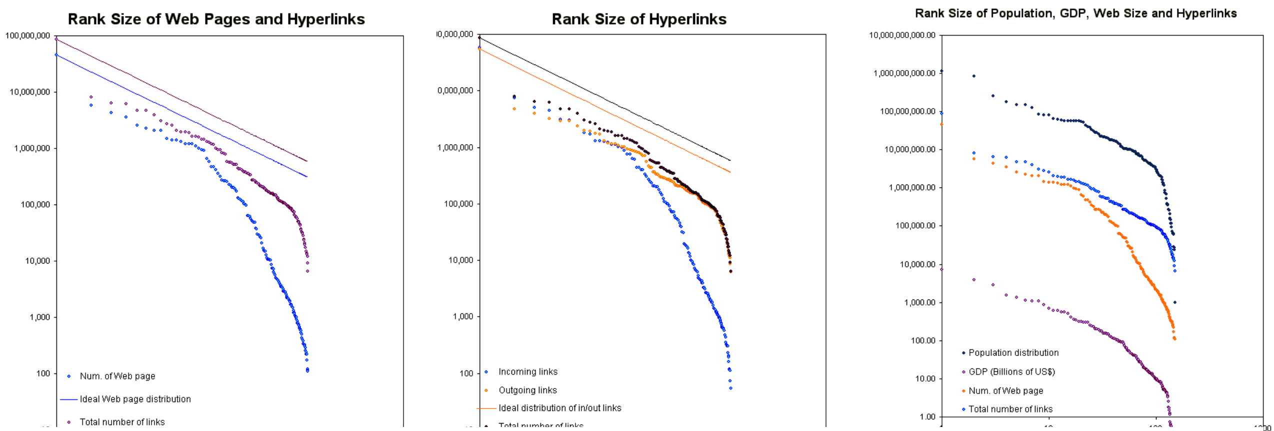


Figure 2: Rank-size data, and power law relationships governing Web size (from left to right): a) Web size and the total number of links; b) incoming, outgoing and total links; c) population, GDP, Web size and total links.

expected (from the rank-size rule, that is) sizes of country in these data is probably as much due to unusual boundaries as to higher growth rates amongst these groups. In contrast, only the first 20 or so domains accord to rank-size when Web page size is examined. This is a classic demonstration of a system undergoing very rapid growth amongst most of its objects with an implication that as one examines successively lower and lower ranks, growth rates would rise inexorably. Of course we have nothing other than Simon's (1957) model to convince us of this, but in terms of more mature systems such as population, the notion is consistent with the data and with our intuition.

Examining the number of links is more problematic. The total and outgoing links conform strongly to rank-size at least for the largest 100 domains measured by these linkages but incoming links is the least like rank-size of any data in our analysis. Again, there is a plausible explanation that outgoing links constitute most of the links in Web pages to date (and maybe forever), and these tend to reflect our perceptions of size while incoming links reflect our ability to link with others. These distributions are quite different and asymmetric in that we tend to know more than proportionately about bigger places than the smaller. This too should change as systems mature. The rank-size relations fitted to these six distributions are shown in the table where we list the intercept, the slope, the correlation squared, and the ratio of the top ranked site's predicted size $P'(1)$ (from the rank-size rule) to its observed value $P(1)$:

Distribution	Intercept log *K*	Slope -*q*	Correlation r^2	*P'(1)/P(1)*
No. Web Pages	21.22	2.91	0.90	35.84
Total Links	18.60	1.60	0.92	1.35
Incoming Links	21.48	2.98	0.89	37.28
Outgoing Links	17.83	1.46	0.91	1.03
GDP	11.98	2.18	0.80	22.67
Population	23.39	2.00	0.72	12.64

These results are statistically rather good but in terms of their actual fit, the evidence of primacy in the top-ranked sites for Web data and for GDP, and the substantial deviations in the short tail for the Web data particularly, reveal that rank-size is only a theoretical ideal which might be attained in the steady state when all domains have been subjected to growth for a long period. To illustrate these points more clearly, we have computed idealised rank-size distributions for each set of data based on $P''(r) = P(1)\,r^{-1}$ where $P''(r)$ is the idealised (pure) value at rank *r* and *P(1)* is the largest observed value in the set. This equation generates a straight line on the log-

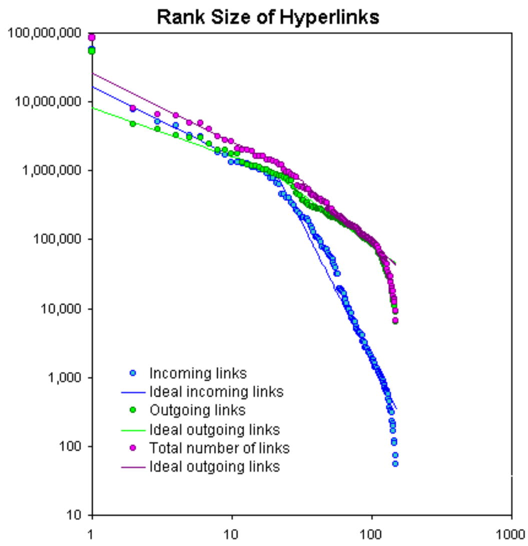


Figure 3. Applying an ideal rank size trend lines with a bent.

log plots and shows how near or far the actual distribution in question is from the steady state. These in fact indicate that the largest sizes do conform well in all cases to rank-size with the shorter tails departing substantially in terms of the slope. For the total Web pages at each site, we have computed two regimes based on the pure rank-size: the first based on the above equation, the second based on $P'''(r > 20) = P''(20)\, r^{-4.25}$ which better mirrors the data in the lower ranges.

Finally we have broken each data set into two ranges by eye and have fitted rank-size relations to each (sample image shown in **Figure 3**). These are shown below

Distribution	Slope $-q_1$ for upper ranks	Correlation r^2 for upper ranks	Slope $-q_2$ for lower ranks	Correlation r^2 for lower ranks	$w_2 q_2 / w_1 q_1$
No. Web Pages	0.88	0.97	4.25	0.98	31.05
Total Links	0.86	0.97	2.07	0.91	15.47
Incoming Links	1.04	0.98	4.49	0.97	26.30
Outgoing Links	0.78	0.97	1.87	0.88	17.29
GDP	1.22	0.99	3.25	0.80	5.65
Population	1.01	0.91	2.80	0.73	1.31

The fifth column shows the weighted ratio between the upper ranks and lower ranks where w_1 and w_2 are the weight of data counted into upper and lower ranks, respectively. These results suggest that there is substantial change still to work itself out within the World Wide Web as the lower ranked sites gradually grow towards the more mature sites at the upper levels of the range. None of this explores how sites change their rank during this process which is yet another matter for future research.

6. Conclusions

Our analysis of the size distribution of global domains and its comparison with the real geography of economic and demographic distributions is the first step in a wider exploration of the shape and structure of cyberspace which promises to enrich our understanding of the information society. The correlations that we found between the size of the Web and population was low although that between the Web and GDP was much higher with an r^2 over 70 percent, confirming our general intuition that the economic development of a domain is all the more important in explaining its size. We anticipate that in time as the global information society matures, the size of the Web will come to reflect the population size of nations much more than it does at present although by then, there will be other specialist Web-like resources which will depend more on the economy than on indicators of demographic size.

Moreover, as the overall rank-size patterns of the Web, its links, and GDP are quite similar, it is perhaps reasonable to conclude that the distribution of Web domains and their links broadly reflects existing economic activity patterns, albeit differences in the distribution pattern of population and Web services. We also expect that Web-based services are carried out at locations remote from places at which these services are initially registered, and we would expect such differences to be reflected in the flows of information between domains – the trade in information between countries. Although our link data contains this, we have not yet been able to explore the patterns contained therein in ways that would confirm this speculation.

The power law relations that we have examined all display the tendency for the number of small events – Web sizes, links, populations, and GDP of small countries – to be less than what the rank-size rule predicts but with Simon’s (1957) model, this can easily be explained by the smaller domains having not yet reached maturity. We did not go as far as to compute growth rates or exponents for every level of rank but we did illustrate the plausibility of the hypothesis that the largest domains approximate the rank-size rule while the smaller domains are growing towards this steady state. The differences in power law that we computed between these two sets confirms this notion. In future work, we will explore these ideas further but to do this, we will require much better data at more than one point in time. This analysis based on a single time-point essentially forms a first step in an interpretation of how Web space is developing. There are many other issues and possibilities that need to be addressed herewith. As well as implementing a time-series analysis, we need to clarify definitions of domains in spatial as well as sectoral terms, and we need to consider suitable spatial and temporal aggregations which affect our analysis.

A major problem is still the definition of the US domain. Super-national level domains such as “com” and “org” require careful estimation as to the extent of their contribution by the US firms and those based in other countries. Some of these large domains were omitted in this study, but their inclusion would significantly alter the value of Web size assigned to the US domain, which in turn, would cause significant changes to the distributions. However it is our belief that the pattern of rank-size would not be markedly altered by such changes, and an essential next step is to see how robust this kind of analysis is to changes in time. Only then we will be in a position to make some tentative predictions as to the future form of cyberspace.

7. References

Abraham, R.H. (1996). Webometry: measuring the complexity of the World Wide Web, (http://thales.vismath.org/webometry/articles/vienna.html).

Adamic, L. (1999). The small world Web, (http://www.parc.xerox.com/istl/groups/iea/www/SmallWorld.html).

Albert, R., Jeong, H. and Barabasi, A-L. (1999). Diameter of the World-Wide Web, *Nature*, vol.**401**, p.130.

AltaVista (1999). AltaVista search engine, (http://www.altavista.com/).

Batty, M. (1993). The geography of cyberspace, *Environment and Planning B: Planning and Design*, vol.**20**(6), pp.615-616.

Berry, B.J.L. and Horton, F.E. (1970). *Geographic Perspectives on Urban Systems*, Prentice-Hall, New Jersey.

Bray, T. (1996). Measuring the Web, *Proceedings for the 5th International World Wide Web Conference*, 6-10 May 1996, Paris, France, (http://www5conf.inria.fr/fich_html/papers/P9/Overview.html).

Carriere, J. and Kazman, R. (1999). WebQuery: searching and visualizing the Web through connectivity, (http://www.cgl.uwaterloo.ca/Projects/Vanish/webquery-1.html).

Carroll, G.R. (1982). National city-size distributions: what do we know after 67 years of research?, *Progress in Human Geography*, vol.**6**, pp.1-43.

Coffman, K.G. and Odlyzko, A. (1998). The Size and growth rate of the Internet, *First Monday*, vol.**3**(10).

Dodge, M. (1998). Journey to the center of the Web, *Telegeography 1999*, Washington DC.

Faloutsos, M., Faloutsos, P. and Faloutsos, C. (1999). On power-law relationships of the Internet topology, *ACM SIGCOMM'99*, Cambridge, MA, pp.251-262.

Gorman, S. (1998). The death of distance but not the end of geography: the Internet as a network. *Working Paper*, Regional Science Association, University of Florida.

Gray, M. (1995). Measuring the growth of the Web, (http://www.mit.edu/people/mkgray/growth/).

Huberman, B.A. and Adamic, L.A. (1999). Growth dynamics of the World-Wide Web, *Nature*, vol.**40**, pp.450-457.

International Monetary Fund (1999). *IMF World Outlook*, IMF.

Kleinberg, J.L. (1997). Authoritative sources in a hyperlinked environment. *IBM Research Report*, RJ10076.

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

Laherrere, J. and Sornette,D.(1998).Stretched exponential distributions in nature and economy, (http:/xxx.lanl.gov/abs/cond-mat/9801293).

Ludwig, G.S. (1996). Virtual reality: a new world for geographic exploration, *EarthWorks*, (http://www.utexas.edu/depts/grg/eworks/wie/ludwig/earthwor.html).

Matrix Information and Directory Services (1999). State of the Internet, Matrix Map Quarterly **603**, (http://www.mids.org/mmq/603/).

Mitchell, W.J. (1999). *E-topia*, The MIT Press, Cambridge, MA.

Murnion, S. and Healey, R.G. (1998). Modelling distance decay effects in Web server information flows, *Geographical Analysis*, vol.**30**(4), pp.285-303.

Network Wizards (1999). Internet domain survey, (http://www.nw.com/zone/WWW/).

Online Computer Library Center (OCLC) (1999). Web characterization project (http://www.oclc.org/).

Pirolli, P., Pitkow, J.E. & Rao, R. (1998). Silk from a sow's ear: extracting usable structures from the Web, *Proceedings for Conference on Human Factors in Computing Systems*, (http://www.acm.org/sigchi/chi96/proceedings/papers/Pirolli_2/pp2.html).

Pitkow, J.E. (1998). Summary of WWW characterizations, *Proceedings for The Seventh International World Wide Web Conference*, 14-18 April 1998, Brisbane, Australia, (http://www7.conf.au/programme/fullpapers/1877/com1877.htm).

Quarterman, J.S. (1999). Internet growth graph, (http://www.mids.org/).

Shiode, N. and Dodge, M. (1998). Visualising the spatial pattern of Internet address space in the United Kingdom. In B.M. Gittings (ed.), *Innovations in GIS 6: Integrating Information Infrastructures with GI Technology*, Taylor&Francis, London, pp.105-118.

Shiode, N. and Dodge, M. (1999). Spatial analysis on the connectivity of the global hyperlink structure. In *Proceedings for GIS Research in UK* (GISRUK'99).

Simon, H.A. (1957). *Models of Man*, John Wiley & Sons, New York.

Sullivan, D. (1999). Search engine watch, (http://www.searchenginewatch.com/reports/sizes.html).

Watts, D.J. and Strogatz, S.H. (1998). Collective dynamics of 'small-world' networks. *Nature*. Vol.**393**, pp.440-442.

Wheeler, D.C. and O'Kelly, M.E. (1999). Network topology and city accessibility of the commercial Internet, *Professional Geographer*, Vol.**51**(3), pp.327-339.

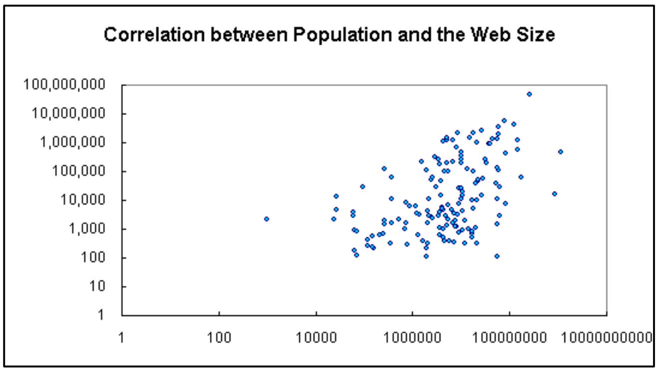
Zipf, G.K. (1949). *Human Behavior and The Principles of Least Effort*, Addison Wesley, Cambridge, MA.

Appendix A: The Full Data Set

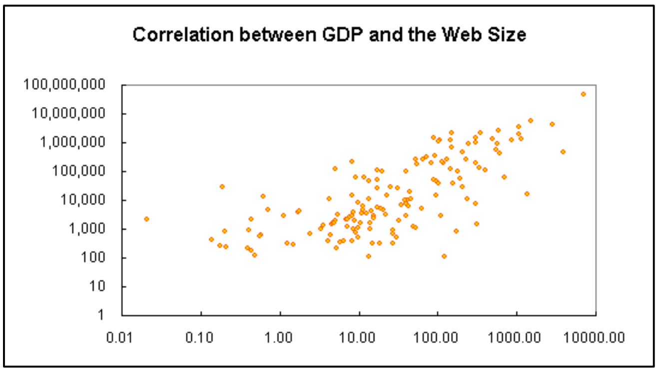
#	Country	Domain	Population	GDP (billions US$)	Web Site Sizes	Incoming Links	Outgoing Links	Total Number of Links
1	Albania	al	1626315	4.077	375	110	149394	149476
2	American Samoa	as	156349	0.396	222	321	148460	148712
3	Andorra	ad	61599	1.116	2793	1786	1920834	1921655
4	Antigua and Barbuda	ag	64794	0.409	871	742	179422	179789
5	Argentina	ar	32712930	304.500	190015	178108	256908	375749
6	Armenia	am	3611700	8.408	3818	2364	106967	108637
7	Australia	au	17661468	347.394	2095633	3009484	1985130	3938057
8	Austria	at	7914127	150.804	660072	751277	867942	1325284
9	Azerbaijan	az	7021178	10.982	3494	1396	100743	101974
10	Bahamas	bs	264175	4.596	1495	1215	95496	96303
11	Bahrain	bh	520653	7.026	2185	1969	28780	30325
12	Barbados	bb	255200	2.464	713	682	157070	157577
13	Belarus	by	10222649	45.791	11397	15974	203379	213927
14	Belgium	be	9967378	204.086	474097	608335	562847	929725
15	Belize	bz	205000	0.588	624	194	23112	23277
16	Benin	bj	4304000	9.893	493	305	31934	32070
17	Bermuda	bm	61220	1.700	3716	3913	61730	63590
18	Bolivia	bo	6420792	20.290	4524	3307	97892	99823
19	Bosnia and Herzegowina	ba	3707000	4.465	632	1001	123422	124039
20	Botswana	bw	1326796	4.364	594	648	59167	59515
21	Brazil	br	150367000	874.490	1198581	1115385	707571	1419328
22	Brunei Darussalam	bn	267800	5.121	2008	1379	44632	45601
23	Bulgaria	bg	8990741	31.060	25610	18049	106245	116986
24	Burkina Faso	bf	9190791	9.182	737	567	32289	32674
25	Cambodia	kh	5816469	6.460	401	602	18888	19274
26	Cameroon	cm	10446409	27.232	926	1024	132663	132989
27	Canada	ca	27408898	598.519	2556128	3134643	2889642	4711362
28	Chile	cl	13599428	146.024	121839	98679	190051	245506
29	China	cn	1136429638	3843.540	468891	199439	202075	354751
30	Colombia	co	27837932	194.264	56175	70752	4668007	4712344
31	Congo	cg	1909248	13.678	109	118	267335	267429
32	Costa Rica	cr	2488749	17.462	50829	77500	139842	190847
33	Cote D'ivoire	ci	10815694	22.869	14827	11528	209686	216264
34	Croatia (Hrvatska)	hr	4511000	19.501	97246	100077	210361	269414
35	Cuba	cu	10743694	14.348	4320	1976	85228	86187
36	Cyprus	cy	725000	9.829	8496	9840	72726	79333
37	Czech Republic	cz	10328017	91.811	358713	441230	251110	517527
38	Denmark	dk	5225689	105.800	1189357	995864	921833	1282025
39	Djibouti	dj	62892	0.444	170	54	58955	58997
40	Dominica	dm	71183	0.200	788	1141	87674	88317
41	Dominican Republic	do	5545741	34.383	7136	8395	162961	167768
42	Ecuador	ec	10740799	44.888	20504	16966	154140	162723
43	Egypt	eg	55163000	235.864	10685	6260	103236	107024
44	El Salvador	sv	4845588	15.535	2890	2688	167207	168919
45	Estonia	ee	1570432	8.149	220819	156117	331001	414766
46	Fiji	fj	715593	4.697	1615	2553	34318	36069
47	Finland	fi	5067620	89.920	1477440	1277732	1093688	1618361
48	France	fr	57526521	1141.601	1384662	1244540	1287696	2064790
49	Georgia	ge	5400841	7.067	2181	2003	96716	97869
50	Germany	de	79364504	1499.874	5760926	5107297	3913423	6432124
51	Ghana	gh	12296081	31.736	1907	2633	35140	36900
52	Greece	gr	10313687	119.533	200666	201923	236204	352358
53	Guatemala	gt	9197351	40.306	7465	6346	51581	55675
54	Guyana	gy	758619	1.487	291	323	12804	13058
55	Holy See (Vatican City)	va	1000	0.021	2107	1209	195522	196640
56	Honduras	hn	4248561	11.187	4682	3961	78150	80570
57	Hong Kong	hk	6686000	113.900	223465	249257	203407	379422
58	Hungary	hu	10323708	64.480	265079	231202	207848	339331
59	Iceland	is	261103	5.036	125834	141512	461785	541806
60	India	in	849638000	1358.980	16620	18741	809657	821241
61	Indonesia	id	179247783	695.034	61010	72207	1158235	1208494
62	Iran	ir	55837163	316.797	1511	1776	89041	90010
63	Ireland	ie	3525719	54.794	172217	197141	839345	946776
64	Israel	il	5123500	82.722	200358	332290	361450	586886
65	Italy	it	57746163	1055.144	2042109	1672011	1731187	2567980
66	Jamaica	jm	2392130	7.773	2522	3112	30229	32504
67	Japan	jp	124451938	2811.027	4291142	7443431	2920306	8024460
68	Jordan	jo	4012000	17.453	5610	4531	82450	85549
69	Kazakhstan	kz	16721113	40.900	10331	4395	10788	13502
70	Kenya	ke	21443636	38.573	10157	17649	33661	44326
71	Korea, Republic of	kr	43663405	500.410	1325365	1828271	898557	1952830
72	Kuwait	kw	2142600	39.703	2926	2664	40623	42452
73	Kyrgyzstan	kg	4451824	8.300	397	473	16823	17161

74	Latvia	lv	2631567	9.060	62736	52721	79765	108607
75	Lebanon	lb	2126325	13.390	10708	13075	41332	49964
76	Liechtenstein	li	27714	0.630	12735	19151	121769	132115
77	Lithuania	lt	3741671	13.480	48447	50535	83938	113038
78	Luxembourg	lu	378400	11.770	62546	71769	170616	219198
79	Macedonia	mk	2055997	1.762	4053	4354	57232	60100
80	Madagascar	mg	7603790	8.978	1939	1323	72981	73818
81	Malaysia	my	18180853	177.544	103451	123586	286836	371280
82	Malta	mt	362977	4.280	10681	13380	149251	155736
83	Mauritius	mu	1168256	11.137	6356	4741	89628	93531
84	Mexico	mx	81249645	611.007	410630	398925	298736	518686
85	Micronesia	fm	118000	0.176	259	438	72631	72992
86	Moldova	md	4360475	8.607	989	1379	157999	159135
87	Monaco	mc	27063	0.721	4617	5329	206497	209522
88	Mongolia	mn	2043400	4.862	1543	939	171934	172360
89	Morocco	ma	26069000	95.438	14828	8509	224195	230475
90	Mozambique	mz	14548400	13.762	998	626	14242	14661
91	Namibia	na	1409920	5.482	3188	3261	153192	155413
92	Nepal	np	17143503	27.265	680	349	82621	82833
93	Netherlands	nl	15184138	299.095	1400750	1226612	1176283	1872236
94	New Zealand	nz	3442500	53.018	264700	398964	310160	559471
95	Nicaragua	ni	3745031	8.192	15129	9888	68176	74510
96	Niger	ne	7248100	5.731	334	299	1004784	1004990
97	Nigeria	ng	55670055	120.614	114	161	441276	441391
98	Norway	no	4286401	103.092	1107890	1100720	1133883	1577234
99	Oman	om	2017591	14.952	330	230	79493	79658
100	Pakistan	pk	84253644	304.174	7420	9432	99234	105773
101	Panama	pa	2562922	15.703	2313	2579	243978	245467
102	Papua New Guinea	pg	3727250	9.733	1053	1114	272317	273152
103	Paraguay	py	4039165	19.016	4978	6773	23700	28320
104	Peru	pe	21998261	93.864	47752	44862	124141	153965
105	Philippines	ph	62868212	203.715	29752	40649	153491	179229
106	Poland	pl	38309226	246.790	941280	882854	2365486	2752733
107	Portugal	pt	9845900	130.311	259240	265625	282969	421907
108	Qatar	qa	369079	10.480	1670	1059	93787	94582
109	Romania	ro	22788969	90.536	50182	54944	84292	118744
110	Russian Federation	ru	148310174	552.555	584276	635463	364514	783502
111	Saint Lucia	lc	148183	0.550	581	576	49691	50078
112	San Marino	sm	23576	0.439	2122	2158	115388	116852
113	Sao Tome And Principe	st	117504	0.136	423	768	431341	431931
114	Saudi Arabia	sa	17119000	175.300	794	633	195952	196385
115	Senegal	sn	6896808	13.837	1682	1795	145966	146988
116	Seychelles	sc	72254	0.480	119	116	273872	273985
117	Singapore	sg	2873800	72.031	321030	303417	294290	516817
118	Slovakia	sk	5318178	40.511	99801	108172	167214	237443
119	Slovenia	si	1990623	16.997	115226	105645	336572	392241
120	South Africa	za	30986920	226.646	270970	437548	271843	570861
121	Spain	es	39141219	559.351	904287	761457	748917	1168910
122	Sri Lanka	lk	17619000	63.510	4906	6577	99960	103704
123	Suriname	sr	354860	1.231	309	205	111813	111930
124	Swaziland	sz	681059	3.337	1040	2247	55313	56937
125	Sweden	se	8692013	152.076	2237539	2350221	1707388	3023385
126	Switzerland	ch	6875364	147.654	1217077	1016403	1027583	1617408
127	Taiwan	tw	20878000	298.500	987654	1278921	703337	1429031
128	Tajikistan	tj	5092603	3.622	1314	1567	22843	24215
129	Tanzania	tz	21733000	18.294	327	876	13880	14608
130	Thailand	th	57760000	405.201	111247	91079	152463	206536
131	Togo	tg	1949493	5.149	217	73	28142	28186
132	Tonga	to	93049	0.192	29149	62087	785745	841328
133	Trinidad And Tobago	tt	1227443	12.342	3501	4292	102605	105806
134	Tunisia	tn	7909555	49.836	1184	913	112983	113434
135	Turkey	tr	56473035	334.941	130324	205634	176272	340697
136	Turkmenistan	tm	3522717	8.269	2791	3260	144939	147542
137	Uganda	ug	16671705	30.618	531	953	58501	59168
138	Ukraine	ua	51801907	102.948	36944	56573	96283	139110
139	United Arab Emirates	ae	862000	42.901	5969	4805	86040	89262
140	United Kingdom	uk	57998400	1064.244	3554483	4497411	3184530	6167812
141	United States	us	258115725	7044.145	45787732	57229750	53917475	86614986
142	Uruguay	uy	3094214	25.511	28432	30477	28068	47901
143	Uzbekistan	uz	19810077	52.344	1149	1309	8624	9115
144	Vanuatu	vu	150165	0.207	241	478	45457	45756
145	Venezuela	ve	20248826	154.581	38043	38768	59248	84680
146	Viet Nam	vn	64375762	111.141	2771	1603	19441	20548
147	Yemen	ye	12301970	27.390	326	315	6239	6446
148	Yugoslavia	yu	10394026	17.000	26419	31582	44396	65735
149	Zambia	zm	7818447	7.239	1193	816	11620	12147
150	Zimbabwe	zw	8687327	22.317	3310	7206	11813	17375

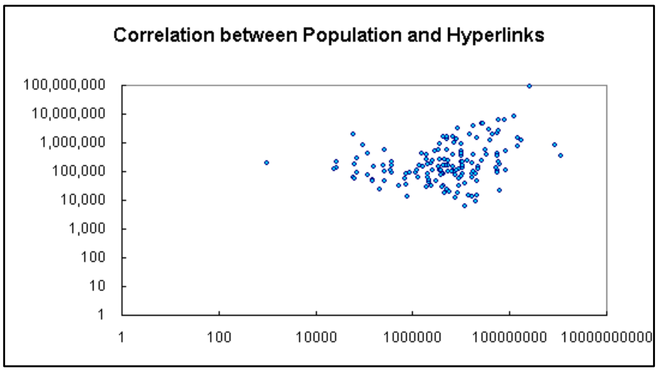
Appendix B: Correlations between Web size, hyperlinks, population and GDP.
Sources: AltaVista (1998), IMF World Outlook (1999).



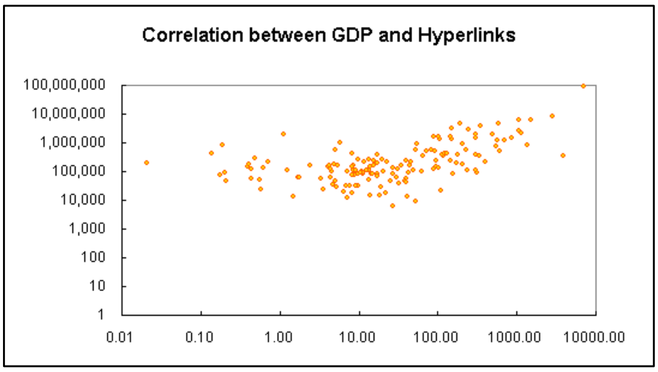
(a) Correlation between population and the Web size (r^2=0.24).



(b) Correlation between GDP and the Web size (r^2=0.74).



(c) Correlation between population and the Web size (r^2=0.09).



(d) Correlation between GDP and the Web size (r^2=0.70).