



Regularities in the Formation and Evolution of Information Cities

DOI:
[10.1007/3-540-45636-8](https://doi.org/10.1007/3-540-45636-8)

[Link to publication record in Manchester Research Explorer](#)

Citation for published version (APA):

Lelis, S., Kavassalis, P., Sairamesh, J., Haridi, S., Holmgren, F., Rafea, M., Hatzistamatiou, A., Tanabe, M. (Ed.), Besselaar, P. V. D. (Ed.), & Ishida, T. (Ed.) (2002). Regularities in the Formation and Evolution of Information Cities. In *Digital Cities II: Computational and Sociological Approaches* (pp. 41-55). (LNCS; Vol. 2362). Springer Nature. <https://doi.org/10.1007/3-540-45636-8>

Published in:
Digital Cities II

Citing this paper

Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights

Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy

If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.



Regularities in the Formation and Evolution of Information Cities¹

Stelios Lelis¹, Petros Kavassalis², Jakka Sairamesh³, Seif Haridi⁴,
Fredrik Holmgren⁴, Mahmoud Rafea⁴, and Antonis Hatzistamatiou¹

¹ Institute for Computer Science, Foundation of Research and Technology - Hellas, ICS-FORTH, Science and Technology Park of Crete, P.O.Box 1385, GR 711 10 Heraklion, Crete, Greece

{slelis, hatzist}@ics.forth.gr

² (Person to contact) Institute for Computer Science, Foundation of Research and Technology - Hellas, ICS-FORTH, Science and Technology Park of Crete, P.O.Box 1385, GR 711 10 Heraklion, Crete, Greece

petros@rpcp.mit.edu

<http://atlantis.uoc.gr>

³ IBM Institute for Advanced Commerce, IBM Thomas J. Watson Research Center, P.O. Box 704, Yorktown Heights, NY 10598

jramesh@us.ibm.com

⁴ Swedish Institute of Computer Science, SICS, Box 1263, SE-16429 Kista, Sweden

{seif, fredrikh, mahmoud}@sics.se

Abstract. In the real world, cities exist because of external economies associated with the geographic concentration of firms within a city. Of course, such a geographic proximity with input providers and consumers, would at first reduce transportation costs. But why cities, information cities, i.e. large agglomerations of people and economic activity emerge in the virtual world? In the Internet, transportation costs are zero. Web sites can easily be reached from anybody and everywhere with no particular cost. In these conditions of equal access distance, one would rather expect a smooth web geography with a relatively even distribution of visitors per site. However, the web economy illustrates strong agglomeration trends with a very small number of web sites capturing a large segment of the web population and the most of virtual economic activity. This paper attempts to provide a sound basis for the dynamics of population concentration in the web under increasing returns.

1 Introduction

Just like physical cities, “information cities” are envisioned to emerge in the future, where each information city will be inhabited by millions (virtual inhabitants) of participants (human or software agents) involved in one or more activities in these cities -- such as buying, selling, chatting, discussing, socializing and collaborating. Evidence of this, is already clear with web portals such as Yahoo!, AOL, Amazon and

¹ This paper summarizes some of the findings of *iCities project*, funded by European Commission (Future and Emerging Technologies, IST-1999-11337), on which we will publish more extensively elsewhere.

others popular web locations (information and/or commercial sites, business-to-business hubs, aggregations of virtual communities) that attract millions of users everyday to conduct one or more activities. Internet industry observers and academics expect many more to emerge in various industrial sectors, regions and so on [1, 2].

However, what Internet observers understand as an increasing trend, i.e. a web geography with disparities between densely populated areas with complex organization (big portals and B2C popular web sites, large B2B marketplaces, virtual communities etc.) and desolate web servers, may appear as a paradox. In the real world, cities exist – according to urban economists and geographers – because of the presence of external economies associated with the geographic concentration of industry within a city [3]. Transportation costs further reinforce the aggregation of firms and workers within large agglomerations. Transportation costs, as commuting costs (i.e. diseconomies of scale) define also the optimal size of the city, thus creating natural limits in cities’ growth process. In the Internet, transportation costs are zero. Why, then, cities would emerge in the virtual world? Given the Internet architecture (no *king* networks, full interoperability) [4], all web sites can be easily reached from everywhere and anybody, with no transportation cost, just at the click of a button. In these conditions of equal access distance, one should rather expect a sort of smooth web geography with a relatively even distribution of the number of visitors per site, across the Internet landscape. Locational dispersion would be easier to understand than concentration and cities’ formation. There would, admittedly, be some irregularities in the density of cyber-population due to the particular attractiveness of certain web sites. It is easy to understand why Disney, CNN, Sony, or Yahoo!, Amazon, e-bay, brand names of the “brick-and-mortar” world or glory pioneers of the “new economy” are able to attract Internet users’ attention. But how did we go from there to large population agglomerations and information cities? Undoubtedly, this is an issue that needs to be elucidated. Especially because it appears that the agglomeration of population over the web illustrates an intriguing empirical regularity.

It is easy to observe that web sites come not only in a large diversity of categories but also in a wide variety of *sizes* (here, the size is defined as a function of the number of visitors). It is probably more difficult to understand why these sizes exhibit a sustained empirical regularity, a power law. In fact, research performed at Xerox Parc with data provided by AOL [5], reveals that the distribution of visitors per Internet site surprisingly follows a universal power law similar to what found in distribution of larger cities in US (and elsewhere) or in income distribution. More precisely, it appears that a small number of sites (about 10%) capture a very large segment of the web population (almost 80%), clear signature of a Zipf-like distribution of the web audience. Very likely, not only audience but also advertising, transactions, e-commerce, all web economic activity are going to be similarly, i.e. non-evenly, distributed across the web. In short, the economic geography of the web space is dramatically bumpy. What may cause concentration in the web and how can we explain these striking empirical data? In this paper we explore potential reasons behind the formation of web agglomeration. We model web population concentration using an agent-based computational model in which aggregate outcomes arise from the bottom-up, through the repeated interaction of autonomous agents. Drawing analogies from Fujita, Krugman and Venables work on the formation of physical cities [6], we especially investigate the contribution of *increasing returns* in the

concentration of virtual social and economic activity in particular places and the emergence of a spatial ordering.

The structure of the article is as follows. In section 2, we document the formation of agglomeration in the web and discuss the striking empirical success of a power law in describing *size* distribution of web sites. We also provide a theoretical approach to deal with this empirical regularity, assuming the presence of some form of increasing returns, with population concentration being self-reinforced. Section 3 introduces this approach in the context of a specific agent-base computational model in which heterogeneous, boundedly rational agents, choose sites to visit (and "stay") in an environment of increasing returns and positive feedbacks. The model reproduces the power law character of the empirical data and provides a simple explanation of this phenomenon mostly based on the mechanisms through which information is transmitted in the Internet. In fact, our results demonstrate that web sites' growth is not only the result of the inherent differences between web sites (crystallized in that might be the "competitive advantage" of a web site), but also a result of some set of cumulative processes, involving *informational increasing returns*, i.e. *word-of-mouth positive feedbacks* and *web surfing path-generated linkages*.

2 The Spatial Order of the Web Economy

The exponential growth of the World Wide Web (W^3) has been accompanied by a global increase in the number of web locations where Internet users (more than 200 millions active Internet users in July 2001 according to Nielsen/NetRatings) spend personal time to get valuable information, socialize and, to a less extent, make transactions and conduct product purchases. Whereas in 1998 there were 2 million of web sites globally, at the end of 2000, the number web sites has increased to 8 million. In addition to this remarkable growth, the diversity of web sites has been considerably reinforced. As a result, Internet users profit from a wider set of location choices, and most likely of a wider range of product versions (and bundles), than those available in the real economy.

However, albeit this extraordinary diversity of web sites, the distribution of visitors across the Internet landscape, illustrates a dramatic unevenness. According to AOL data collected and interpreted by Xerox Internet Ecologies Project, the distribution of larger web sites is surprisingly well described by a power law [5] – the pattern is so universal that applies both in the case of all sites and sites in specific categories (adult sites, educational sites). As Adamic and Huberman [ibid.] explain and Fig. 1 and Table 1 prove, the existence of a power law implies that a small number of sites command the traffic of a large segment of the web population².

² Formally speaking, the number of web sites whose visits is exactly P , is proportional to P to some negative power (P^p). We consider that for a particular site, the number of (unique) visitors it receives in one day, week or month, illustrates site's popularity and, of course, relates to its *size*.

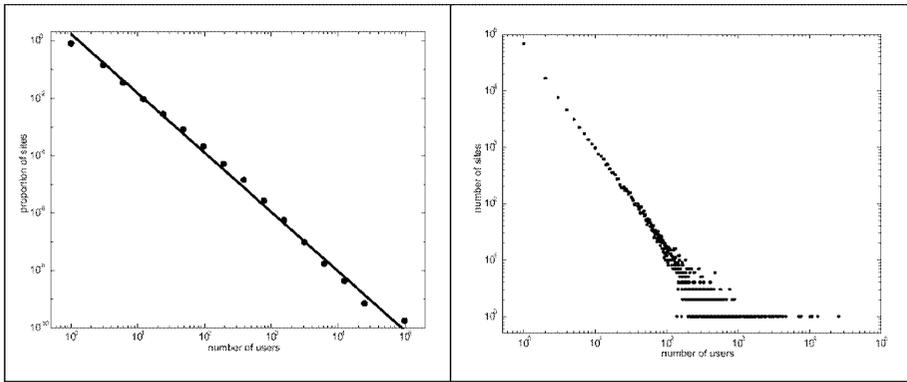


Fig. 1. Occurrence of sites by popularity according to Xerox Internet Ecologies Project (vertical axis: proportion of sites/number of sites, horizontal axe: number of users)

Table 1. Distribution of user volume among web sites according to Xerox Internet Ecologies Project

% sites	% users volume		
	all sites	adult sites	educational sites
0.1	32.36	1.4	2.81
1	55.63	15.83	23.76
5	74.81	41.75	59.50
10	82.26	59.29	74.48
50	94.92	90.76	96.88

We reached exactly the same conclusion by examining a different set of data provided to us by PC Data Online³. In this sample as well, distribution of unique visits among web sites is fairly described by a power law where the top 10% sites capture almost the 60% of the attention of the sampled users. More interesting, web sites ranked in top-high positions remain relatively stable in their rank in a percentage that approaches, during the examined period, the 50%. While relative ranking stability is a clear characteristic of the sample, data provide also clear indications for new sites succeeding to reach high-top ranking positions. To give an example, the panel of 1000 top-high positions is constantly renewed (i.e. weekly) in a percentage close to 15%. However, falls from the high-top positions and rises in top ranks of new sites are

³ PC Data Online (now at <http://www.netscoreonline.com>) had monitored the behavior of a sample of 10,000 Internet users, in both free of access web sites and proprietary content online providers. Data we have used refer to the period February to September 2000. Adamic and Huberman [5] obtained the distribution of web users among web sites from access logs of a subset of AOL users active December 1, 1997. The subset comprised 60,000 web users accessing 120,000 web sites.

rarely dramatic. They happen rather incrementally, and certainly within a longer period of time.

What is interesting in observing this parameter, the number of visitors or the popularity of a site? Of course, the number of unique visitors has *de facto* become an indicator for the market success of a web site (a short of an index for observing aggregate demand). But it also captures how spatial population accumulation happens across the web. That signification we retain here. At least, except for *eyes-balls* advertisers, the number of visitors does not say so much about the profitability and the competitive positioning of a particular web site in its industry segment. Surely, it is more valuable for urban planners of the virtual space who want to understand the population aggregation/segregation trends and for web designers who design the appropriate infrastructure allowing a web site to serve its customers. In short, we perceive the number of visitors for a web site as the equivalent of the *city size* in the physical world. Likewise, the city metaphor has some relevance to the matter in hand. Exactly as in the urban ecology paradigm [7], the size of web sites, as they evolve towards information cities, together with density and population heterogeneity, might become important to study: i) the nature of these cities, ii) the features of social and economic relationships found in them and, iii) the social norms that will emerge in virtual urban environments⁴.

Furthermore, exactly as in the size distribution of physical cities [8]⁵, we can see in the power law describing web sites' *sizes* distribution the existence of some hidden principle at work, proof of order arising from self-organization. It is likely that the Internet exhibits sustained empirical regularities that are as striking and consistent as those of the urban and regional economy. Maybe, one can learn a lot by studying models of economic geography in which there is concentration of activity in particular locations. What economic geography does principally have to say to the Internet? The response holds in just a word: *increasing returns* [9]. In a spatial economy with increasing returns, competition between locations to attract industries, involves mechanisms of positive feedback systematically working to reinforce the winner-places (by amplifying an initial advantage) and challenge less quickly growing locations. Arthur [10], Fujita, Krugman and Venables [6] and Krugman [8], all see geographical concentration in the real world as the result not necessarily of

⁴ There is an essential difference in the notion of size in the real and the virtual world. Physical cities' habitants live only in one location while *infohabitants* may easily live different "lives" in many virtual spaces. However, in reality, the number of these "lives" is rather limited or, putting it in another way, the *portfolio of choice of an infohabitant* is not very large in size. What is a portfolio of choice? The unique sites an Internet user visits during a specific time period (usually companies that collect these statistics work on weekly or monthly basis) is often referred as repertoire or portfolio. An Internet user may visit a certain web site several times during a week/month but it would count only once in its portfolio. According to Nielsen/NetRatings, the average portfolio size of European Internet users in April 2001 ranged from 13 (Sweden) to 26 (Germany), while the average portfolio size of US Internet users has been 20.

⁵ A rank size rule applies to the distribution of physical cities sizes (i.e. the population of a city is inversely proportional to its rank), which comes to approach a Zipf-distribution. More precisely, there seems to be something close to a log-linear relationship between rank and population (i.e. a straight line) with the slope of this approximately linear relationship to be almost -1.

inherent (quality-based) differences among locations, “but of some set of cumulative processes, necessarily involving some form of increasing returns, whereby geographical concentration can be self-reinforcing” [6]. In a similar manner, if the web locational system generates so much agglomeration (i.e. population accumulation) in particular places (portals, auction sites, specialized content providers), it is reasonable to suppose that something is working behind the surface to prevent dispersion and keep Internet consumers more and more congregated in a few places. Something that shapes the dynamics of competition and produces imperfect competition-like results⁶ in an economy with certainly lower transactions, compared with the real world economies, and low market friction [11]. We would argue that, exactly as in the spatial economy, a credible story about web spatial development should focus on the role of increasing returns. Our sense is that increasing returns in the virtual space mostly arise from *informational feedbacks*: interactions among economic agents in the virtual space are found on information flows, generated through information dissemination, information collection and processing, and information exchange processes. To give an example, in a chaotic space such as the web, information gathering and much of individual learning about web sites can be viewed as a sampling or “polling” process of the experience of other Internet users, i.e. previous visitors in these web sites (i.e. *increasing returns from word-of-mouth feedbacks*)⁷.

To proceed further we need to model informational increasing returns as emerging from the behaviour of individual agents. We need, in other words, to design a robust model able to give deeper insights to web concentration phenomenon while integrating some of the above mentioned informational increasing returns. The other point is that the model should also generate something that looks as a power law. One way to model power law phenomena would be to assume a stochastic process of growth that can produce a large range of sizes whose upper part will follow a power law distribution⁸. However in this paper, following Krugman’s suggestion, we are more tempted to explore a model *involving random networks of interaction among agents* rather than random growth. “The randomness”, Krugman explains, “that creates the power law may not involve random growth but random ‘connections’ in space. For example, imagine port cities that serve the interior along a transport network formed with random connections among transport nodes, with the direction of the preferred connections reflecting accidents either of history or geography. Alternatively, we could suppose that the connections lie in some abstract space of industry linkages...”[6]. In the next section, we will describe a computational agent-based model that relies on two “random” networks: one network transmits *word-of-mouth information* among Internet consumers; another one (*the underlying network*) defines relational paths between web sites that make navigation very hierarchical. Both structures are allowing for *random connections* to be generating and both “carry” increasing returns.

⁶ As Adamic and Huberman notice, “sites are rewarded by relative performance than absolute performance” [5].

⁷ Not strangely, word-of-mouth or *viral marketing* in the professional jargon, is largely recognized as a powerful arm for efficient web marketing.

⁸ This is the approach taken by Simon [12] to model real cities’ rank-size rule and by Adamic and Huberman [5] to model web markets.

3 Power Law-Based Competitive Dynamics of Web Sites under Informational Increasing Returns: A Computational Model with Agents

A model in which heterogeneous agents select web sites to visit and “stay”⁹ is described in this section. Their behavior is quite complex. To effectively navigate in the web, they may explore by their own means, opportunities offered by this immense space (i.e. web sites that can satisfy their tastes). Sometimes, given the high uncertainty of that environment, they collect information that may help them to better navigate from asking advice from previous visitors, sampling their experience – which sites they have visited and appreciated. Apparently, the central cognitive issues raised in this model are these of *exploration along a transport network* and *information contagion* within local social networks. The goal of this modeling work is to understand how an aggregate distribution of web sites as a function of their size may emerge from the behavior of individual agents.

3.1 Behavioral Foundations

We start with two stylized facts. First, suppose an Internet user surfing on the Net. Which sites will he visit? He may start from interesting sites that he has visited in the past, one or more times (bookmarks file). Or, he may discover new ones, through media information and advertising or by using a search engine to help him. Alternatively, he may ask advice from friends; or a friend of him will show up and tell him, “Oh, have you seen this site, it's really good!”. Obviously, which sites an Internet user will decide to visit merely depends on which sites other Internet users have “polled” or sampled and decided to visit and stay. As a result, web sites’ popularity appears to be self-reinforcing and determined by a sort of historical process in which references from previous users, friends, and acquaintances, drive sites growth¹⁰. Second, suppose an Internet user visiting a particular web site. It is plausible that while navigating within the site, he will be tempted to click on related or suggested links that drive to other web sites. That means, which sites an Internet user decides to visit also depends on the number of other sites pointing to it. The more *in-links* a site has, the more users will visit it. Recent work in the area of web navigation patterns has shown that although it is not always the case, pages with high in-degree will also have high-ranking positions [13].

The first stylized fact describes a process that is similar to the individual consumers’ reliance on word-of-mouth communication when they decide to buy a product with performance unknown ex-ante (in the sense that the real performance of

⁹ In the sense that they include these web sites into their *portfolio of frequently visited sites* (see footnote 4).

¹⁰ Evidence to this story is provided by a recent story published by The New York Times (22.03.200). *Marketers Try Infecting the Internet: Overwhelmingly, Web users are relying on friends and family for new site recommendations, the report said. Fifty-seven percent say word of mouth or viral marketing is their main source of information about new sites. So instead of relying on banner advertisements and expensive marketing campaigns to promote a site or service, companies are looking to their customers as a source of publicity.*

the product, and its suitability to purchaser's particular needs, will be proved only after use). Under these circumstances, the potential purchaser will probably try to reduce this uncertainty by asking previous purchasers about how they have evaluated the different products in competition. However, as demonstrated by Arthur and Lane [14], Banerjee and Fundenberg [15], Ellison and Fudenberg [16], reliance on this sort of easily obtained information (through sampling or "pooling") may affect the aggregate outcome by introducing an informational feedback which can cause, in certain cases, market shares to become self-reinforcing. Arthur and Lane [14] use the term *information contagion* for this informationally generated linkage between a product's prevalence and its likelihood of being purchased. We adopt the same view about how information is transmitted among Internet users and how this introduces an information feedback into the process of web sites competing for market share. In our model, the sites that users learn about depend on which sites others users have visited and "stayed", so Internet-users, through word-of-mouth feedbacks, are likely to learn more about popular sites than unpopular ones, sites with few previous visitors. The second stylized fact relates somehow to the *linkages* story that explained the formation of physical cities [6]. The idea is that firms and people are finding benefits to be near other firms because they want to explore potential backward and forwards linkages. Almost the same process takes place in the web where sites find benefits to be "near" sites that have large in and out degrees (surfing-path generated linkages). In our model, web sites are positioned along a transport network, defining relational paths between web sites that make navigation very hierarchical.

3.2 Computational Implementation

The model is computationally implemented using object-oriented programming. Two populations of software agents have been created: a population of Internet users making "purchasing" decisions among several products-web sites, and a population of web sites to which we attribute different performance characteristics. The Agent Based Environment gets functionality from *Mozart*, a distributed software architecture developed by SICS¹¹.

3.3 Set-Up of the Model

Fig. 2 represents our web model economy, consisting of M Internet consumer agents and N web sites agents (M and N increase linear over time). Each agent's internal parameters are printed inside its eclipse.

A1: Initial conditions. Initially ($t=0$), Internet consumer agent i , $i \in (1.. M_0)$, select randomly lp_0 web sites, among N_0 , to create an initial version of its *portfolio of choices* (therefore portfolio, a subset of their bookmarks folder). Web site agent j , $j \in (1.. N_0)$, has performance characteristic r_j that determines its performance in practice (that is, its "intrinsic value"). A random uniform distribution describes sites' performance at $t=0$; *I, performance interval*, illustrates the interval between upper and

¹¹ See Mozart Consortium: The Mozart Programming System (<http://www.mozart-oz.org/>).

lower performance values. We suppose that these performance characteristics do not change over the course of the locational choice process described by the model.

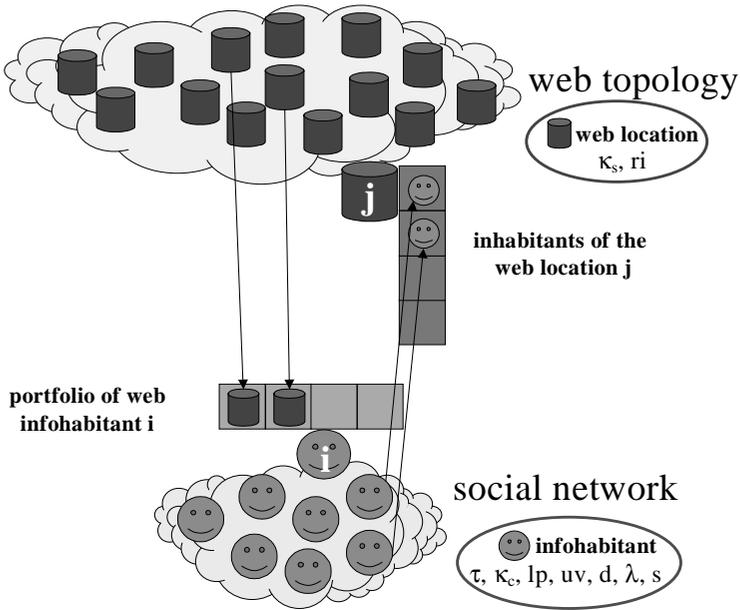


Fig. 2. Graphical representation of a simulated web economy

A2: Agents' activation. Internet consumer agents are activated at each time step t . Agents are activated exactly once during a single period (we suppose uniform activation, at least at this implementation of the model). When an agent is activated, starts a new surfing session (i.e. navigation across different web sites) that ends when time step t expires. Naturally, agent i starts a new surfing session (at time t) from visiting a subset uv of the web sites included in the portfolio it has composed during the previous time step ($t-1$).

A3: Portfolio's composition and update. The surfing behavior of consumer agents is crystallized in their portfolios. In fact, frequently visited sites are selectively organized into a portfolio of choices, which is different from one agent to another. The length of that portfolio, l_p , remains constant over the model course (at least at this implementation of the model). Agents' portfolio is regularly updated (at each time step t) as follows. From time step t_0 ($t=0$), individual agents have randomly selected l_{p_0} sites to create an initial version of their portfolio of favorite sites. As the process evolves, agents learn about new sites via a word-of-mouth-based information gathering mechanism (A4) and through personal exploration of the web (A5). Agents evaluate sites they learn about according to a utility function U , and include in their portfolio the sites with the highest perceived utility (i.e., high-ranked sites after a performance check). In addition, we assume agent i to be relatively loyal to its portfolio: a new site will be included in it, only if agent i perceives a high utility for

that site over a longer period of time (i.e., we require that consumer agents retain a site and include it in their portfolio if they perceive a higher utility after number λ of positive checks during a period of s iterations; formally we call λ the number of times agent i calculates that a particular site j gives a higher utility before including it in his portfolio).

A4: Learning through word-of-mouth. Internet consumer agents that have been activated, decide first whether they will ask other users to collect information about “interesting sites”. Only a fraction α of consumer agents ask other agents (friends and acquaintances) to propose favorite sites. So, at each time step t , agents α randomly selected from the whole population, randomly sample τ other agents, where τ is a stopping rule (a decision rule for terminating sampling, which is supposed the same for all agents, in this model implementation). Of course, consumer agents interact with other agents near to them to exchange information about web sites they have already visited and appreciated. We suppose a *small world* structure¹² describing the dynamics of this interaction; κ the number of initial neighbors in the local social network and β the probability of edge rewiring in this network. Each sampled individual proposes to the agent who poses the question, one (number 1) web site, randomly selected from his portfolio. Agents then visit these sites, evaluate them and select the sites that will be included in their portfolio.

A5: Exploration along hierarchical navigation paths. Consumer agents surf from one web site to another along hyperlinks. More precisely, at each time step t , agent i visits a number of d sites following the *out-links* of the sites it has visited earlier during the iteration (uv_i plus τ_i). Following Adamic [18], the *underlying network* of links connecting web sites is considered as a *small world* graph, where vertexes define web sites, and edges define links among them¹³. Agents then visit these sites, evaluate them and select the sites that will be included in their portfolio.

A6: Use’s evaluation method. We assume that the utility U_{ij} agent i perceives from visiting site j is $rj + e$, where e represents agent-specific idiosyncratic shocks, which could capture either variations in sites’ download time (for example, how fast, today, one can access this or that site), or variations in agent’s willingness to explore the site.

¹² It is largely accepted that social networks are described by a *small world* structure, that is: even when two persons are not friends or do not have common friends, there is a sort chain of acquaintances that connect these two persons. Watts [17] define small world networks according to two basic properties: i) The average shortest path from any two users is small and, ii) the clustering coefficient (that is a measure of the extend with which users band together) is large, compared to the clustering coefficient of random graphs. He proposes three different network models that appear to embody the defining characteristics of the small world structure. In our model, we use Watts’ *β -model* in order to generate a graph representing a social network transmitting word-of-mouth information. The algorithm starts with a highly ordered graph, a perfect 1-lattice, in which each vertex has precisely κ neighbors, and randomly rewires the edges of the lattice with probability β . In the resulting graph, which illustrates the characteristics of small world graphs, vertices have an average number of neighbors equal to κ with small variance.

¹³ We again use Watts’ *β -model* to generate a graph representing the network defining navigation paths (β and k take different values in the social and underlying network).

We suppose that e are individually independent distributions (i.d.d.) over time and across agents and sites. For simplicity we assume that e has a normal distribution with mean 0 and standard deviation σ_{ob} .

3.4 Aggregate Dynamics: Distribution of Web Sites' Sizes

In this sub-section we describe typical realizations of this model. Results are still exploratory and presenting initial observations, rather than describing definitive conclusions and sensing trends.

3.4.1 Evolution in Time of Web Sites' Sizes. From Fig. 3 it can be seen that our model reproduces a power law distribution that appears as a straight line in log-log coordinates (parameter assumptions: $\alpha=0.2$, $\tau=1$, $d=2$, $lp=15$).

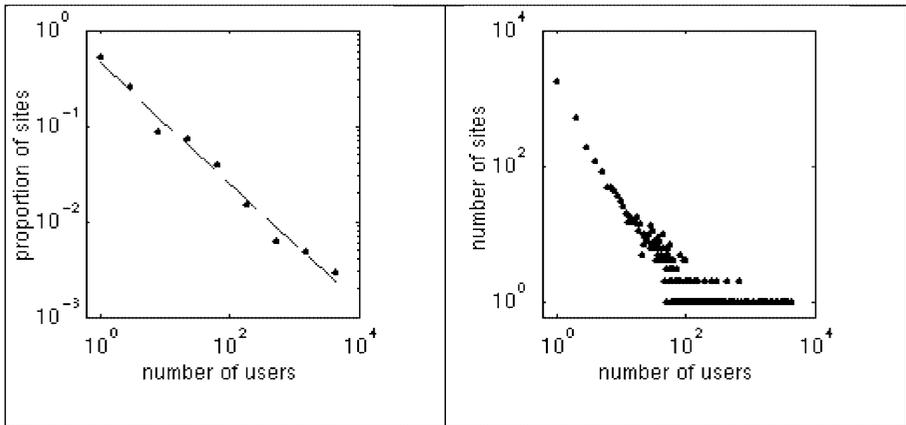


Fig. 3. Distribution of web sites by size, iCities project results (vertical axis: proportion of sites/number of sites, horizontal axis: number of users)

The above figure is a simple run of a model of a web economy initially consisting of 10,000 Internet users and 5,000 web sites. Web sites size distribution at the iteration number 50,000 ($t=50,000$) is described by a power law with exponent $\rho=0.63$. Interestingly, numerical results are pretty close to these provided by Xerox Internet Ecologies Project with AOL data.

Table 2. Distribution of user volume among sites: iCities and Xerox Internet Ecologies Project results

% sites	% users volume	
	all sites/Model results	all sites/Xerox results
0.1	13.98	32.36
1	63.54	55.63
5	84.31	74.81
10	90.92	82.26
50	98.44	94.92

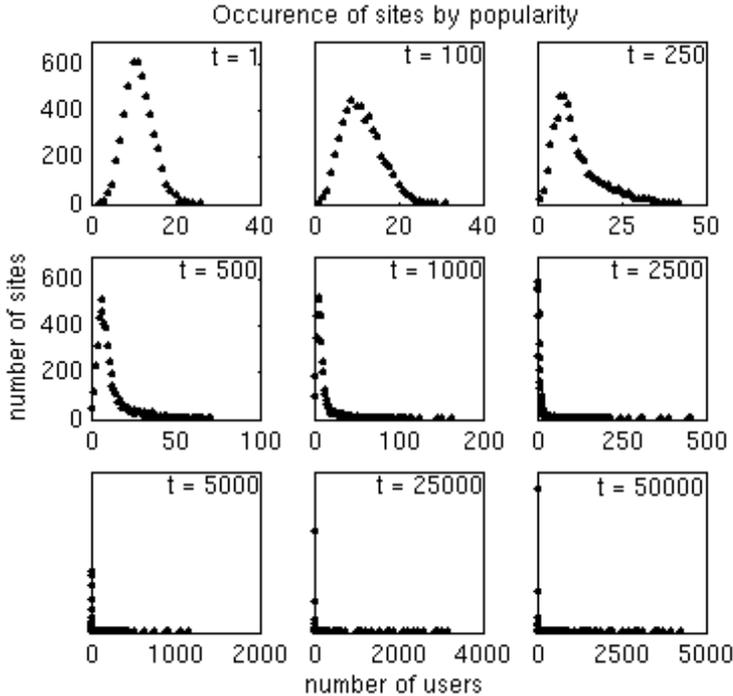


Fig. 4. Histogram “number of sites/number of users”: evolution in time

We started the run ($t=1$) with a non-even distribution of Internet consumer agents across the 5,000 web locations (Fig. 4). As agents are being activated and new consumers and sites entering the game ($M_{\text{final}}=19,880$, $N_{\text{final}}=9,909$), the pick of the histogram gets closer and closer to 1 ($t=2500$). Already at time $t=2,500$, a group of sites appears to gain large proportions of unique visits. Hereafter, most of the sites progressively lose their appeal while few sites continue to gain more and more visitors. As a result, the shape of the curve changes and progressively takes a L-like form. This is an evidence of the transition from a normal distribution with a small mean to a power law distribution. Regarding the exponent of the power law distribution, it stabilizes close to 0.63 after time step $t=25,000$. Notice also another property inherent in the model dynamics. We have established that a few locations would become big by exploring word-of-mouth information diffusion and by navigational hierarchies, but, according to our experiments, neither of these effects (case $a=0$ or $d=0$) would alone produce a power law distribution result.

3.4.2 Unpredictable Evolution Paths, Variable Growth Rates, and Relative Performance Awards. Our results give also some indications about eventual patterns of self-organization that should be further investigated to get definitive conclusions. First, we should notice that the “game” is very uncertain at the early stages of the

model course (up to $t=2,500$). Of course, (early) success begets success, but there is always the possibility for sites that had a slow start to accelerate, during later stages, their rate of growth and finally rank with the more popular web sites. The same applies for new sites, entering the competition at later stages. Necessary condition for this is that they are *strongly connected* with already popular sites. In addition, small differences in initial conditions seem to have large effect on the final outcome. Second, web locations with the largest sizes seem to belong to small groups of co-evolving locations (Fig. 5). This may not be strange given the small world topology of the *underlying network* but it is interesting to observe that the web space organizes itself into a structure composed of groups of sites of similar size with a characteristic distance between them.

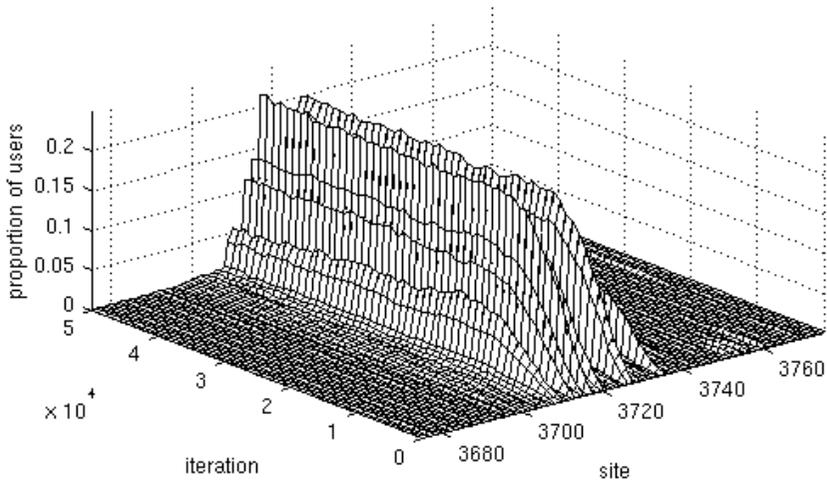


Fig. 5. The evolution in time of the stronger group with the most popular sites

Third, a very interesting result comes from observing growth rates of different web locations. For locations that succeeded to rank into the upper positions, size seems to fluctuate considerably at the early stages of the process but it progressively stabilizes. On the contrary, sites with very limited success, or no success, remain with highly fluctuating growth rates over the whole model course. Is that also a sign of self-organization? Fig. 6 illustrates the differences in growth rates between the winner location and the site ranked into the 125th size position.

Finally, no strong correlation seems to exist between web sites' ranking and the performance characteristics. Figure 7 confirms that in many cases, sites with comparatively low performance have been attracting more users than sites with high performance values.

Notice that the word-of-mouth mechanism displays all the increasing returns properties of informational feedbacks: not only does it privilege sites that became quickly established and early well known, but it also acts as a powerful mechanism of exclusion for others sites with a relatively good performance.

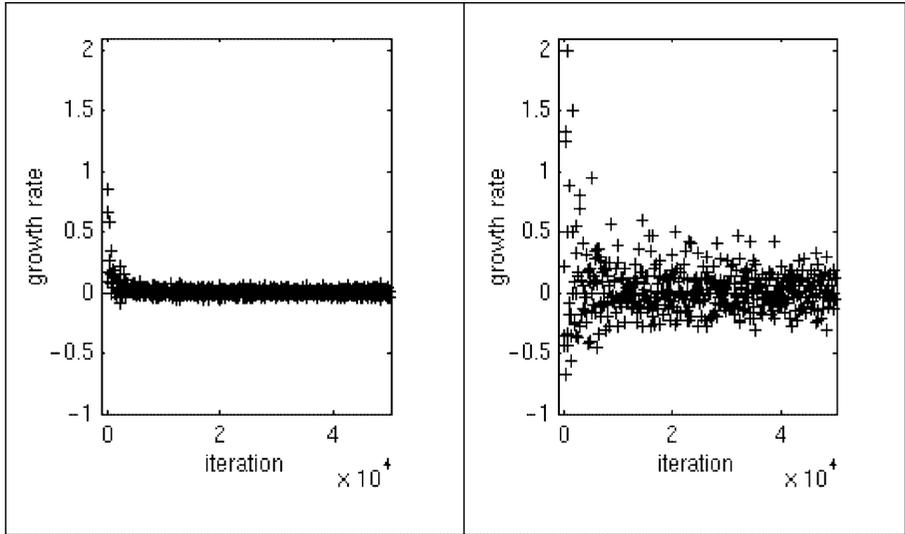


Fig. 6. Differences in evolution of growth rates for sites ranked into 1st and 125th positions

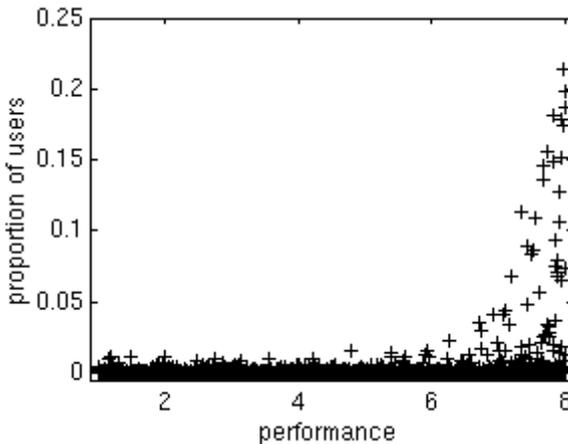


Fig. 7. Size versus performance ranks

3.4.3 Robustness of the Results and Issues under Further Investigation. The results we presented in this paper are highly exploratory and require further investigation in two directions. First, we need to better understand the behavior of the power law distribution, especially the dependence of its exponent ρ on basic model assumptions. Second, we need to experiment with other underlying network topologies, closer to the Internet reality. Broder et al. [19] have recently revealed a more detailed and subtle picture of the web architecture, much richer from a small world-like connectivity: “in a sense the web is much like a complicated organism, in

which the local structure at a microscopic scale looks very regular like a biological cell, but the global structure exhibits interesting morphological structure (body and limbs) that are not obviously evident in the local structure (i.e. bow-tie graph)".

References

1. Sairamesh, J., Nikolaou, C., Kavassalis, P., Haridi, S.: Building an Information City over the Internet: Design Principles and Architecture. Digital Communities Conference (2001)
2. Isida, T., Isbister, K. (eds): Digital Cities: Technologies Experiences and Future Perspectives. Lecture Notes in Computer Science, Vol. 1765, Springer-Verlag, Berlin Heidelberg New York (2000)
3. Henderson, J. (1998) Urban development: Theory, Facts and Illusion. Oxford UP.
4. Gillet, S.: The Self-Governing Internet. Coordination by Design. MIT ITC Working Paper
5. Adamic, L., Huberman, B. The nature of Markets in the World Wide Web. QJEC 1 (2000) 5-12
6. Fujita, M., Krugman, P., Venables. A.: The Spatial Economy. Cities, Regions and International Trade. MIT Press (1999)
7. Kleniewski, N.: Cities, Change and Conflict. Wadsworth Publishing Company (1997)
8. Krugman, P.: The Self-Organizing Economy. Blackwell Publishers (1996)
9. Arthur, B.: Increasing Returns and Path Dependence in the Economy. University of Michigan Press (1994)
10. Arthur, B.: Industry Location Patterns and the Importance of History. In: Arthur, B.: Increasing Returns and Path Dependence in the Economy. University of Michigan Press (1994) 49-67
11. Bailey, J.: Intermediation and Electronic Markets: Aggregation and Pricing in Internet Commerce. Ph.D. Thesis. MIT (1998)
12. Simon, H.: On a class of Skew Distribution Functions. *Biometrika* 42 (1955) 425-440
13. Brin, S., Page, L.: The anatomy of a large scale hypertextual web search engine. 7th WWW Conference (1998)
14. Arthur, B., Lane, D.: Information Contagion. Increasing Returns and Path Dependence in the Economy. In: Arthur, B.: Increasing Returns and Path Dependence in the Economy. University of Michigan Press (1994) 69-97
15. Banerjee, A., Fundenberg, D.: Word of Mouth Learning. Working Paper. Available at <http://fudenberg.fas.harvard.edu/mouth.htm>.
16. Ellison, G., Fundenberg, D. Word of Mouth Communication and Social Learning. *QJE* 110 (1995) 612-643
17. Watts, D.: Small Worlds. The Dynamics of Networks Between Order and Randomness. Princeton University Press (1999)
18. Adamic, L.: The small world web. 3rd European Conference On Digital Libraries (1999)
19. Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A., Wiener, J.: Graph Structure in the Web. 9th WWW Conference (2000)